DNA sequence analysis of T cell receptors

Joseph, Ansamma K.

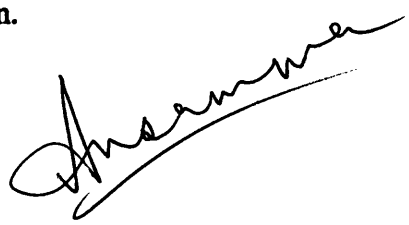
Award date:
1996

Awarding institution:
University of Bath

DNA SEQUENCE ANALYSIS OF

T CELL RECEPTORS

submitted by Ansamma K. Joseph
for the degree of PhD
of the University of Bath
1996



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

DNA sequence analysis and molecular modelling of T cell receptor (TCR) variable regions (Vα and Vβ genes) has been carried out to explore the relationship between TCR sequence and structure.

Analysis of the Vα and Vβ genes revealed a much greater variability between different TCR sequences than is seen when different antibody variable regions are compared.

Different variable region sequences recognising the same MHC-peptide specificity generally showed a higher degree of similarity, both within the framework and CDR regions, as compared to those sequences recognising different MHC-peptide specificities.

Molecular models of six TCR variable regions from DR restricted clones were generated in an attempt to understand the relationship between MHC restriction and TCR sequence and structure. This part of the analysis has highlighted the fact that even small differences in sequence can produce significant differences in surface topology.

The combined sequence and modelling analysis has led to the following conclusions. Closely similar MHC molecules containing identical peptides do not necessarily elicit a *highly* restricted T cell receptor population. This surprising result leads to the notion that therapeutic strategies in certain autoimmune diseases such as rheumatoid arthritis might better be targetted to the MHC peptide cavities than to a

diverse TCR population.

When peptides are presented in the context of different MHCs, the expected selection of different TCR sequences and topologies is observed.

Dedicated to my parents

Acknowledgements

I would first like to thank my supervisor Tony Rees for his generous support and encouragement throughout the entire course of this work. I would also like to thank all past and present members of the Rees group especially Graham Elliott, Jeffrey Gyi, Jonathan Milner, Robert Griest, Phil Williams, Alison Jones and Sue Phillips for their helpful suggestions and useful advice.

A special thanks to Nick and Steve for their help with modelling, Toula and Tassos for advice on MOLSCRIPT and Peter Keller for general computing programs.

I am also grateful to David Sansom at Bath Institute of Rheumatic Diseases for his help and advice in the initial cloning experiments and Paul Moss from the Institute of Molecular Medicine, Oxford for advice on Anchored PCR.

I would also like to thank my aunt Dr A. Ninan for giving me this opportunity to experience the wild west and my uncle Ninan and cousins Golda, Silva and Shola for their infinite generosity and patience throughout the entire course of this project. A very special thanks to Ravi Acharya and Vasanta for their support and encouragement at various crisis points (and there were a great many of them!!).

Finally I would like to thank Sonali, Liat and Ollie for moral support and providing lighter moments during this project which has made this 'experience of a lifetime' all the more enjoyable.

Abbreviations

Chemicals

BSA	bovine serum albumin
DEPC	diethyl pyrocarbonate
DMSO	dimethyl sulphoxide
DTT	dithiothreitol
EDTA	ethylene diamine tetra acetic acid
IPTG	isopropyl thiogalactoside
MOPS	3-(N-morpholino) propane sulphonic acid
PBS	phosphate buffered saline
PMSF	phenyl methyl sulphonyl fluoride
TCA	trichloroacetic acid
TEMED	N,N,N',N'-tetramethylethylenediamine

General

Å	angstrom unit, 1 Å = 0.1 nm
AbM	Antibody Modelling algorithm
ACS	antibody combining site
APC	absolutely positively conserved
C	constant
CDR	complementarity determining region
cDNA	complementary deoxyribonucleic acid
Da	Dalton
dCTP	deoxy cytosine triphosphate
dGTP	deoxy guanosine triphosphate
dNTP	deoxyribonucleoside triphosphates
DNA	deoxyribonucleic acid
GCG	Genetics Computer Group

HEL	hen egg lysozyme
HPLC	high pressure liquid chromatography
Ig	immunoglobin
L-broth	Luria broth
mAb	monoclonal antibody
MHC	major histocompatibility complex
MPI	minimum pairwise identity
OD	optical density
oligo(dT)	oligodeoxythymidylate
polyA	polyadenylated
RF	replicative form
RNA	ribonucleic acid
SDM	site directed mutagenesis
ss	single stranded
TCR	T cell receptor
TAP	transporters associated with antigen processing
Tm	melting temperature
tRNA	transfer RNA
UV	ultraviolet
V	variable
WT	wild type
w/v	ratio of weight to volume (g/ml)
w/w	ratio of weight to weight

Contents

List of figures

List of Tables

CHAPTER 1

Introduction

1.1 Immune Recognition

The immune system in mammals has evolved primarily to recognise and destroy the large variety of potential pathogens which an individual may encounter. The ability to specifically recognise molecules on pathogens is a basic feature of the adaptive immune response and is mediated by T and B cells. B cells recognise antigens in their native conformation either free in solution, on membranes or on the surfaces of cells, using surface immunoglobulin (Ig) as a specific antigen receptor. Binding of the antigen is followed by its processing and presentation in the context of MHC molecules on the surface of the B cell and interaction with T cells. This induces expansion of the B cell clones specific for that antigen and differentiation into plasma cells capable of producing and secreting large amounts of soluble antibody of the same specificity. The secreted antibody triggers processes such as complement activation or macrophage stimulation ultimately leading to removal of the antigen. T cells recognise antigen only in association with these specific major histocompatibility molecules (MHC). This phenomenon is known as MHC restriction. The recognition of protein antigens by T cells involves a highly specific interaction between a T' cell antigen receptor and a peptide epitope presented in the groove of a specific MHC molecule on the surface of B cells or antigen presenting cells (APCs). In addition, accessory molecules are required to trigger the biological

response (CD4, CD8, LFA-1, LFA-3).

An understanding of the molecular recognition events that occur in antigen -antibody and TCR-peptide-MHC complexes is important for the study of the immune system and is essential for the development of therapies for diseases that involve the immune system.

1.2 Biochemical characteristics of the T cell receptor and its accessory molecules

1.2.1 The TCR αβ heterodimer

The TCR is a member of the immunoglobulin supergene family. It is a multichain molecular complex with variable and conserved components (Figure 1.1). It consists of a hypervariable disulphide linked heterodimer (TCR-α,β or TCR-γ,δ) that is noncovalently linked to a five chain complex (CD3-γ,δ,ε ; ζζ or ζη), CD4 or CD8 coreceptors and CD45 (reviewed by Janeway, 1992). Assembly of all subunits is required for efficient surface expression. The variable chains of the receptor (Ti-α, β, γ, δ) are entirely responsible for the recognition of peptide-MHC complexes on the surface of antigen presenting cells (Schwartz, 1985). Each of these chains consist of variable (V) and constant (C) domains stabilised by intrachain disulphide bonds, a short hinge like segment, transmembrane (TM) and intracytoplasmic portions. The Ti portion of the molecule is polymorphic and carries the T cell epitope.

The invariant CD3 complex consists of four non-covalently associated

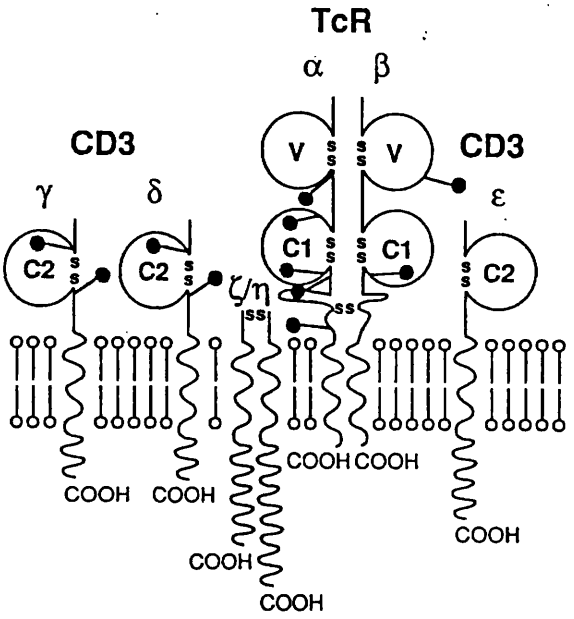


Figure 1.1 Schematic diagram of the T cell receptor-CD3 complex.

homologous chains γ, δ, 2ε. Another chain, CD3 ω or TRAP (T-cell receptor associated protein) associates with the CD3 complex during the assembly of the TCR in the ER. It contains large cytoplasmic domains ranging from 44-81 amino acids. Both Ti and CD3 subunits are non covalently associated with a ζ dimer. The ζ chain is a transmembrane protein with an extracellular domain, a single transmembrane domain and a cytoplasmic domain of 113 amino acids (Weissman *et al.*, 1988; Baniyash *et al.*, 1989). While the ζ chain exists primarily as a homodimer a small % of TCRs in certain T cell clones contain heterodimers of ζ complexed to the η chain (an alternatively spliced product of the ζ chain) [Baniyash *et al.*, 1988; Orloff *et al.*, 1989; Jin *et al.*, 1990] or the ν chain of the multi-subunit high affinity IgE receptor (FcεR1) [Orloff *et al.*, 1990]. The ν subunit of FcεR1 shares sequence homology in both transmembrane and cytoplasmic regions with the TCR ζ chain. These chains are non-glycosylated and exist as disulphide linked dimers. The ζ chain plays an important role in mediating TCR signal transduction (Chan *et al.*, 1994) The two signal transducing modules, CD3 and ζ chain subunits have common peptide sequences within their cytoplasmic domains termed immuno-receptor tyrosine-based activation motifs (ITAMs) that are responsible for transducing signalling events.

The TCR has no intrinsic protein tyrosine kinase (PTK) activity. However three cytoplasmic PTKs, $p59^{fyn}$, $p56^{lck}$ and ZAP-70 have been implicated in proximal TCR signalling. These cytoplasmic PTKs fall into two distinct categories which differ structurally: $p59^{fyn}$ and $p56^{lck}$ are members of the lymphocyte specific *src*-family (Samelson *et al.*, 1990, 1992; Abraham *et al.*, 1991) while ZAP-70 is a

member of the *syk*-family (Chan *et al.*, 1992). The co-receptor molecules CD4 and CD8 are transmembrane glycoproteins and each is associated with a molecule of the T cell specific tyrosine kinase $p56^{lck}$. This may account for their ability to influence the TCR complex-mediated signaling.

Another critical component of the signal transduction machinery of lymphocytes is CD45, which has intrinsic tyrosine phosphatase activity. It also plays an important role in antigen-stimulated proliferation of T cells and in thymic development.

1.2.2 The TCR γδ heterodimer

The γδ receptors are present on a small percentage of peripheral T cells (0.5 -10%) in humans (Brenner *et al.*, 1986; Bank *et al.*, 1986; Lanier *et al.*, 1987) and on a majority of Thy-1^+ dendritic epidermal cells (Stingl *et al.*, 1987*a* and *b*; Koning *et al.*, 1987). This heterodimer appears earlier (around day 13) than the α:β dimer (day 17 or 18) during ontogeny of TCR positive cells. It is found either as a disulphide linked or a non disulphide linked heterodimer (Lanier *et al.*, 1987; Krangel *et al.*, 1987 and Hochstenbach *et al.*, 1988). The nature of the ligands and precise function of this receptor remain elusive. The molecular nature of γδ T cell recognition is fundamentally different from that of αβ T cells (reviewed by Raulet, 1994). Most γδ T cells develop normally in mice lacking MHC Class I and Class II molecules (Bigby *et al.*, 1993; Correa *et al.*, 1992; Grusby *et al.*, 1993) which questions the role of these MHC molecules in the normal function of gamma-delta T cells. Analyses of gamma-delta T cell hybridomas specific for

classical MHC Class II (IE) and for TL Class I molecules indicate that the topology of the interaction of the γδ T cell receptor with the MHC is distinct from that of αβ T cells and that MHC bound peptides do not influence TCR binding (Schild *et al.*, 1994). This may suggest that some TCRs bind to MHC molecules from the side rather than from the top, away from the peptide binding groove. Comparative measurements of the lengths of the third complementarity determining regions (CDR3) of α, β, γ, δ and TCR chains and Ig heavy and light chains suggest the possibility that γδ TCRs may be more Ig like in their antigen recognition properties (Rock *et al.*, 1994). They recognize native antigens without a requirement for degradation and presentation by antigen presenting cells (Schild *et al.*, 1994).

The great diversity of γδ TCRs is created mainly by junctional diversification and they appear to play a role in protection against microbial infections. Direct evidence that γδ T cells provide immunity to infectious diseases was demonstrated in mice deficient for αβ T cells (Hiromatsu *et al.*, 1992; Mombaerts *et al.*, 1993; Tsuji *et al.*, 1994; Kaufmann *et al.*, 1994). Mice deficient for both αβ and γδ T cells were found to be extremely sensitive to disease caused by the intracellular bacterium *Listeria monocytogenes*, whereas mice deficient for either αβ or γδ T cells were found to be similarly resistant to primary infections (Ladel *et al.*, 1994). In another study T cell deficient mice were shown to be unable to resist liver infection with *Plasmodium yoelli* (Tsuji *et al.*, 1994). Stimulated γδ cells have been shown to secrete lymphokines and to lyse target cells.

1.3 TCR gene organisation

The TCR genes are rearranged and expressed in the thymus during the earliest stages of T-cell differentiation. β, γ and δ chain genes are rearranged and transcribed first followed by the α chain genes. They are split into arrays of interchangeable coding segments scattered over large tracts of chromosomal DNA. Specific V, D (in the case of β and δ) and J segments recombine in a developing T cell to produce a contiguous V(D)J exon. This exon is spliced together with the C region at the level of RNA.

1.3.1 The TCRα gene organisation

The α chain gene is located on chromosome 14 at 14q11 and is unique in that the δ TCR gene complex lies within Jα and Vα (Caccia *et al.*, 1985; Collins *et al.*, 1985). About 70 Vα and over 70 Jα gene segments are spread over 100 kb of DNA. (Figure 1.2). The Jα segments occupy approximately 85 kb and lie about 4 kb upstream of the single Cα segment (Yoshikai *et al.*, 1985). Based on nucleotide homology of greater than 75%, the human Vα gene segments may be grouped into 32 families with at least 15 subfamilies containing a single member. Largest subfamilies Vα1 and Vα2 contain 9 and 7 members respectively. The Cα gene is split into four exons. The external part of the Cα region is encoded by two exons. The transmembrane segment and short cytoplasmic tail are encoded by a third exon with the 3' untranslated region of the mRNA encoded by the fourth Cα exon. A recent study has shown that up to one-third of mature T-cells express two Vα

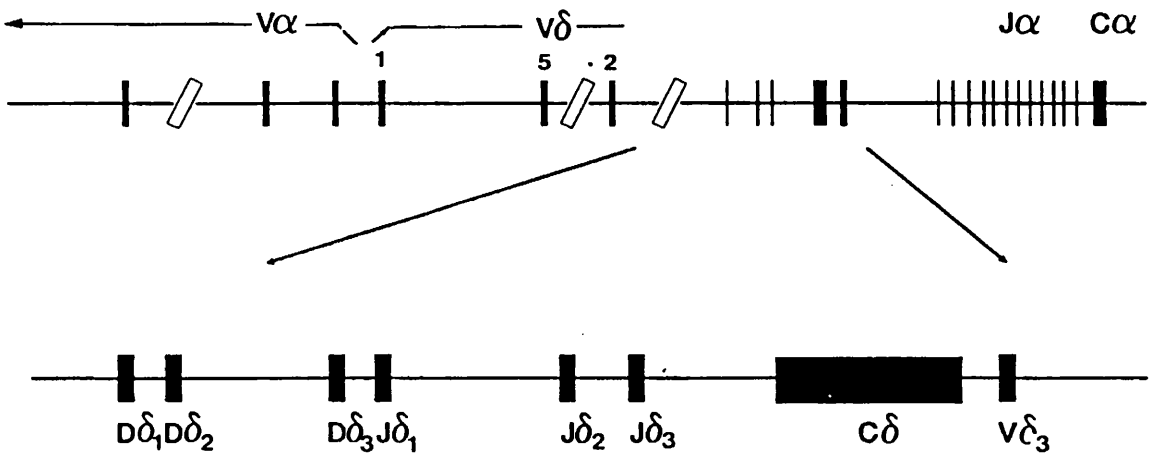


Figure 1.2 A schematic representation of the TCR α/δ gene complex on chromosome 14 (taken from Moss *et al.*, 1992).

chains as part of two functional and independent T-cell receptors. Cells belonging to this dual TCR subset may be specific for a broader range of antigens than cells with a single receptor (Padovan, 1995).

1.3.2 The TCR β gene organisation

The TCR β chain gene complex spans almost 685 kb on chromosome 7 at 7q35 (Rowen *et al.*, 1996). It has 57 different functional Vβ genes, 2 Dβ genes, 13 Jβ and 2 Cβ genes (Figure 1.3). The two Cβ gene segments are separated by about 8 kb. They are highly homologous differing in only 6 amino acids, which are not likely to mediate any functional differences between Cβ1 and Cβ2 (Toyonaga *et al.*, 1985). Both genes have the same transcriptional orientation and consist of 4 exons with introns located at identical positions in each gene. Most of the extracellular C domain is encoded by the first two exons. The third exon codes for major part of TM portion and the fourth exon encodes a small cytoplasmic tail of 5-6 amino acids and the 3' untranslated region. Upstream of each Cβ segment there is a cluster of seven J segments (six functional and one pseudogene) and one D segment. The Dβ1 and Dβ2 segments are located about 600 base pairs 5' to the corresponding Jβ cluster. The Jβ gene segments encode 15-17 amino acids of the Vβ region. The Jβ1 cluster contains 6 functional segments and the Jβ2 cluster contains 7 functional genes. Vβ gene segments are divided into 24 subfamilies, the majority with 1 or 2 members.

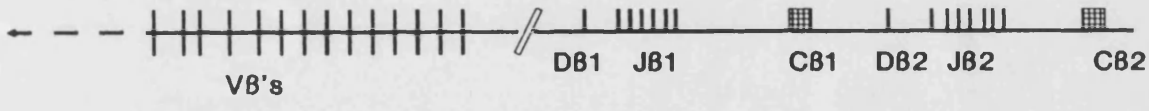


Figure 1.3 The TCR β chain complex (taken from Moss *et al.*, 1992).

1.3.3 TCRγ gene organisation

The Vγ gene is localised in the short arm of chromosome 7 (Murre *et al.*, 1986). Vγ gene organisation is complex. There are 14 Vγ segments, 8 of which are functional, located upstream of 2 sets of Jγ segments (Figure 1.4). The TCRγ locus contains four C segments one of which (Cγ3) is a pseudogene. The Cγ1 gene is composed of 3 exons. The first exon encodes the extracellular domain, the second encodes most of the connecting peptide and the third exon encodes the remainder of the connecting peptide, transmembrane region, 12 amino acids, cytoplasmic tail and 3' untranslated region. A cysteine residue is encoded by the second exon of Cγ1 which is involved in interchain disulphide bridging. Cγ2 has a similar organisation except that Cγ2 alleles contain either duplicated or triplicated second exon segments. Allele 2bc is present in 68% of the population while the allele 2abc is the minority form (Li *et al.*, 1988).These 3 Cγ2 exons do not encode a cysteine residue due to which some γδ receptors do not have an interchain disulphide bond. There are no germline Dγ segments. The human Jγ gene segments are more extensive with three identified upstream of Cγ1 and 2 upstream of Cγ2.

1.3.4 TCRδ gene organisation

The TCR δ locus is unique in that many of the TCRδ coding segments are located entirely within the V and J segments of the α gene (Hata *et al.*, 1987; Chien *et al.*, 1987; Loh *et al.*, 1987). This unusual location results in the deletion of Cδ prior to rearrangement of Vα to Jα using specific 5' and 3' δ-deleting elements

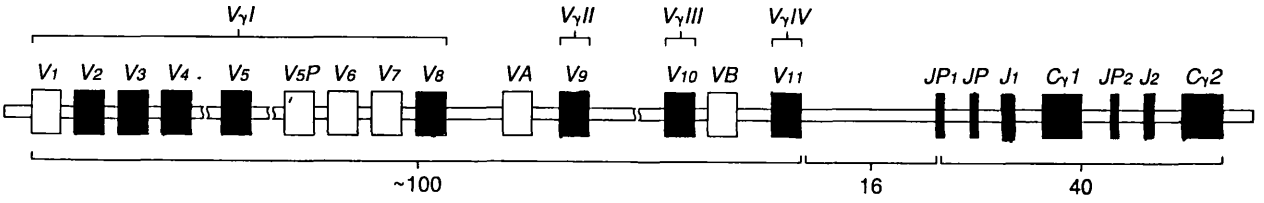


Figure 1.4 Genomic organisation of the human TCR γ locus (taken from Male *et al.*, 1996). The V genes, some of which are pseudogenes (unshaded) are clustered over about 100 kb and are divided into four subgroups.

termed δRec and ΨJα (Hockett *et al.*, 1988). At least 10 distinct human Vα segments rearrange at the Jδ locus which leads to scrambling of the two V gene repertoires. Genevée *et al* (1994) have shown that in addition to a wide junctional region diversity characteristic of the δ chain there was potential for extended combinatorial diversity due to recruitment of at least 1/3rd of the Vα gene segment subfamily repertoire in the δ locus recombination process (n=13) to 3 Dδ and at least three or four Jδ gene segments.

1.4 The TCR repertoire

The available TCR repertoire is determined by a number of factors including inherited germline gene segment repertoire, preferential rearrangement of certain TCR V-gene subfamilies, positive and negative selection in the thymus as well as peripheral tolerance induction or idiotype interactions (as a result of which all the expressed receptors will not be available for antigenic recognition). T cell repertoire is constantly changing as new T cells emigrate from the thymus.

In man, differences in genomic organisation of the TCR region may have important functional implications. In certain mice strains, germline V region deletions and allelic variation of some V regions are seen. Another genetic factor determining peripheral repertoire is the T cell selection during maturation. In mice, deletion specific for Vβ elements also occur and is determined by endogenous superantigens such as Mls.

Another major influence on human and murine circulating T cell repertoire is the effect of environmental antigens including SAgs acting as mitogens for

subpopulations of T cells. A novel mechanism which influences T cell repertoire in a Vβ dependent manner, independent of SAg or TCR structure has been described (Kay *et al.*, 1994). A bi-allelic locus TCR BV13S2 gene was identified where allelic differences predominate in the non-coding region including transitions, transversions and frameshift deletions. Expressed protein is non-polymorphic at this locus. TCR BV 13S2 genotype influenced the circulating levels of Vβ 13.2 CD4 T cells but did not affect TCR expression and function. Altered TCR repertoire is generated by germline non-coding polymorphism and this mechanism also provides an alternative explanation to SAg for Vβ bias observed in certain autoimmune diseases (Paliard *et al.*, 1991 and Sumida *et al.*, 1992).

1.4.1 Mechanisms creating diversity in the TCR repertoire

The mechanisms by which diversity is created in the TCR repertoire are similar to those used by B cells, with one exception. The large number of gene segments allows combinatorial diversity to generate many transcriptional permutations, thus increasing receptor diversity. Diversity is further increased by N region addition of non-germline encoded nucleotides at the V(D)J and Jα-Cα junctions. The template independent DNA polymerase TdT has long been implicated as the enzyme responsible for the addition of N nucleotides. It is expressed almost exclusively in immature thymocytes and pro-B cells. It is capable of adding N nucleotides to artificial recombination substrates *in vitro* (Landau et al., 1987; Kallenbach et al., 1992).

Another source of variation in some TCRs is palindromic P-nucleotide addition (Lafaille *et al.*, 1989). Although an exonuclease is generally invoked for the loss of nucleotides, both variable nucleotide loss and gain of P nucleotides may be caused by asymmetric scission of closed hairpin termini prior to joining (Lieber *et al.*, 1992). TCR genes, however, do not undergo somatic hypermutation, unlike antibodies, thus avoiding the potential for generating auto-reactive specificities outside the selective environment of the thymus.

The '12-23' rule governs TCR gene rearrangements. Early *et al* (1980) proposed that a gene segment with a 12 base pair spacer can only recombine with a gene segment containing a 23 base pair spacer. All TCR V genes have 2-turn recognition signals at their 3' flanking regions whereas Jα, Jβ and Jγ gene segments are associated with 1-turn 5' recognition sequence. The Dβ1 and Dβ2 segments utilise a 1-turn 5' recognition and 2-turn 3' recognition sequence. Multiple rearrangement combinations are possible at the TCRβ locus. V-D-J and V-J rearrangements are allowed by the '12-23' rule.

In human peripheral blood there is a marked non-random usage of individual V region families; several families Vβ4, Vβ5, Vβ6, Vβ12 and Vβ13 make up the majority of the repertoire. Different individuals of random HLA type have a broadly similar profile of Vβ repertoire although PCR analysis has shown that identical twins are more similar to each other than to the outbred population implying a genetic component in the repertoire (Moss *et al.*, 1992).

The Jβ repertoire shows significant variation in the usage of Jβ segments in TCR transcripts. The Cβ2 complex is utilised in about 65% of all transcripts. Within

each Cβ complex there is variable usage of different J segments that does not correlate with relative positions in the germline complex. Analysis of sterile JβCβ transcripts shows a similar Jβ profile implying that no obvious selective pressure acts to dictate the Jβ profile.

Very little information is available on the murine or human Vα repertoires because of the large number of Vα and Jα segments and the few monoclonal antibodies available to characterise them. The usage of Vα segments by lymphocytes in human peripheral blood was studied by Moss *et al.*, 1992. A similar profile is seen in different individuals although differences of the order of 3-fold are seen for some Vα segments such as Vα4 and Vα11. The repertoire is indistinguishable between identical twins, again suggesting a predominant genetic influence on the Vα genes.

TCR gene rearrangement generates considerable junctional diversity due to 'imprecision of the joining process'. This mechanism for the generation of diversity is important for TCR γ genes which have a relatively restricted germline gene repertoire.

1.4.2 Analysis of the TCR repertoire

T cell repertoire analysis helps to understand the cellular and molecular laws that govern the immune response in both physiological and pathological conditions.

Methods used to analyse TCR repertoire essentially fall into two groups, use of monoclonal antibodies to Vα or Vβ segments and recombinant techniques

which attempt to quantitate levels of TCR mRNA.

Monoclonal antibodies have been widely applied as they have the advantage of speed, ease of use, reproducibility and unlimited source. It is also the only technique that measures cell membrane expression. However antibodies cannot generate information on J and junctional sequences and currently cover only a partial portion of the V region repertoire.

Recombinant techniques can be broadly divided into Rnase protection or PCR- based techniques. The former has the advantage of not needing nucleic acid amplification (Baccala *et al.*, 1991).

PCR based methods that have been commonly used for the amplification of cDNA reverse transcribed from TCR RNA include 'quantitative' PCR (Choi *et al.*, 1989; Sottini *et al.*, 1991), inverse PCR (Uematsu *et al.*, 1991), anchored PCR (Loh *et al.*, 1989 ; Rosenberg *et al.*, 1991), the use of consensus primers (Broeren *et al.*, 1991) and sequence analysis within the VDJ region of the PCR products (Uematsu *et al.*, 1991).

V region family specific PCR is the most rapid method but the exponential nature of this reaction allows for substantial variation in efficiency to occur depending on the set of primers used and may account for some of the discrepant results. As each V region is amplified in separate reactions it is difficult to make quantitative comparisons of the % usage of different V regions so that it will be necessary to normalise TCR mRNA in samples being compared. Often *in vitro* expansion of T cell populations is necessary prior to the use of this technique which

introduces an additional source of bias. Another limitation of this technique is that novel V regions locking the primer specificity cannot be detected.

A rapid method for cloning TCR V regions was described by Uematsu *et al.*, 1991. In this method the cDNA is circularised and two adjacent primers in an inward-outward fashion are used for PCR. It requires second strand synthesis, a blunt ligation and is further complicated by intermolecular ligation. Degenerate oligos have been used for the amplification of antigen receptors (Larrick *et al.*, 1989), DNA binding proteins (Lai and Lemke, 1991), olfactory receptors (Buck and Axel, 1991). Danska *et al.*, 1990 designed degenerate primers to take advantage of a highly conserved region of the TCR α and β chain V regions centered around the WY residues at amino acids 34 and 35 (Kabat numbering, 1992) and used them for the amplification of mouse Vα chains from murine T cell clones. The TGGTA core is highly conserved in a number of human Vα and Vβ genes. Moonka and Loh (1993) designed a shorter consensus primer for the amplification of TCR Vα and Vβ genes. The degeneracy is greater (23328 vs 216) and equimolar concentration ratios were used for each mixed base rather than biasing the mixture according to the frequency of occurence. The consensus primer has low affinity for variable regions which lack the TGGTA core namely Vα13.1, Vα19.1 and Vαw23.1. This method is highly specific, easy to use, sensitive, reliable and inexpensive in comparison to other routinely used methods.

A modification of PCR, called Anchor PCR, which creates artificial 3' ends on all cDNA prior to PCR to allow hybridisation of a single complementary 3' primer has been used. Anchor PCR allows amplification of all transcripts by the

same set of primers. It is quantitative and not dependent on known V region sequence data. This technique ensures that no bias is introduced as in the case of the other techniques.

PCR based analysis has been less reproducible than use of antibodies. Cellular TCR mRNA levels may be distorted by the state of activation of T cell subsets. Some groups expand T cells in vitro or activate whole populations with anti CD3 antibodies to ensure that sufficient RNA is available for detection. This may introduce an additional bias. When allied with sequencing it gives data on the whole TCR transcript.

1.5 Antigen recognition by T cells

1.5.1 MHC restriction of T cells

The TCR recognises denatured or processed antigens only in association with MHC antigens, a phenomenon known as MHC restriction. On the basis of their MHC restriction, T cells can be separated into two main functional subsets, regulatory (T_H/T_S) and effector cytotoxic (T_C) T cells. Helper T cells (T_H) recognise antigen in association with MHC Class II determinants whereas T_C cells use MHC Class I as restriction elements. Helper cells express CD4 (62 kd glycoprotein) while cytotoxic cells express CD8 (76 kd glycoprotein). These cell membrane glycoproteins are reciprocally expressed on two subsets of mature T cells in the blood and lymphoid tissues ($CD4^+CD8^-$ or $CD4^-CD8^+$) while a large number of immature T cells in the thymus express both molecules $CD4^+CD8^+$ while most immature cells are $CD4^-CD8^-$.

The coreceptor molecules CD4 and CD8 are members of the Ig gene superfamily (Doyle *et al.*, 1987 and Norment *et al.*, 1988). The extracellular domains of CD4 and CD8 interact specifically with conserved non-polymorphic sites on Class II ($\beta 2$) and Class I MHC ($\alpha 3$) molecules respectively (Konig *et al.*, 1992; Salter *et al.*, 1990; Cammarota et al., 1992). These interactions result in increasing intercellular adhesion and enhanced stimulation of T cells. They exhibit significantly different structural features.

The D1/D2 domains (Ryu *et al.*, 1990 and Wang *et al.*, 1990) and D3/D4 domains (Brady *et al.*, 1993) of the soluble CD4 molecule have been determined by X-ray crystallography. The N-terminal Ig like domain of CD4 exists as a monomer while in CD8 it forms an Fv like dimer. Four Ig like domains D1-D4 are present in the extracellular regions of CD4 while the extracellular region of CD8 consists of a single Ig-like domain connected to a transmembrane segment by a 'stalk' region of approximately 48 amino acids. There are six residues in the CDR2 like loop in both CD4 and CD8 which are not present in Ig V domain. The functional significance of these structural features is unclear but Fleury *et al* (1991) have shown that the extended CDR2 like loop in CD4 is unlikely to be involved in interactions with Class II molecules as deletion of this loop fails to impair CD4 Class II MHC cell adhesion.

The ability to distinguish "self" from "nonself" has been mapped to the major histocompatibility complex (MHC). The human MHC is known as the human leucocyte antigen (HLA) system and is located on the short arm of chromosome 6 at 6p21.3. Klein (1986) classified MHC molecules on the basis of structure and

function. Three classes of molecules, denoted I, II and III were identified in the human MHC.

There are three commonly recognised Class I loci HLA-A, B and C encoding classical transplantation antigens and three other loci HLA-E, F and G known to encode functional molecules. Two pseudogenes HLA-H and HLA-J have also been recognised. The Class II loci are located in the HLA-D region which is organised into DP, DQ and DR families with the exception of one alpha (DZ_α) and one beta chain gene DO_β. There is only one HLA-DR α chain gene (HLA-DRA), 5 HLA-DR β chain genes (HLA-DRB 1-5) and 4 pseudogenes (HLA-DRB 6-9). Class III genes encode several components of the classical complement pathway, factor B of the alternate pathway and an enzyme involved in steroid biosynthesis, 21-hydroxylase.

1.5.2 Peptide-MHC interactions

One of the main features of MHC molecules is their extreme polymorphism and over one hundred structurally distinct Class I and II molecules have been identified. About six different Class I and II molecules are expressed on the cells of a fully heterozygous individual. The expression of multiple isoforms as well as the structural variety available within a species ensures that at least one peptide derived from any pathogen is encountered. As each isoform is capable of binding to a wide array of distinct peptide antigens, a tremendous diversity of peptide-MHC complexes can be formed, allowing the immune system to respond specifically to a variety of pathogens. In addition to broad peptide specificity and the ability to

present each peptide as an antigenically unique complex, the MHC molecules must be able to retain their ligands on the cell surface at an effective free peptide concentration of zero, long enough for T cell recognition to occur.

The three dimensional structural analysis of several MHC molecules reveals an elegant solution to the multiple constraints of flexible but stable binding, with the ligand exposed for recognition.

1.5.2.1 Class I MHC- peptide interactions

Class I molecules are expressed on almost all nucleated cells although the level of expression does vary between cell types.

The three dimensional structures of three human (HLA-A2, HLA-Aw68 and HLA-B27)[Madden *et al.*, 1991; Madden *et al.*, 1992; Silver *et al.*, 1992; Madden *et al.*, 1993] and two murine (H-2K^b and H-2D^b)[Zhang *et al.*, 1992; Fremont *et al.*, 1992; Young *et al.*, 1994] class I molecules, in complex with single peptides and with mixtures of endogenous peptides, have been determined by X-ray crystallography. The overall domain arrangement and backbone structure of these complexes are very similar.

They are comprised of a glycosylated polypeptide chain (α heavy chain, Mr = 44k) in close non-covalent association with the invariant β_2 microglobulin (β light chain, Mr =12k). The heavy chain consists of three extracellular domains α1, α2, and α3, a transmembrane region and a cytoplasmic tail. α1 and α2 are structurally similar and together form an eight stranded antiparallel β-sheet spanned by two long antiparallel helical regions while α3 has an Ig fold. The Class I β chain,

β_2 microglobulin is a single domain which has an Ig fold similar to the $\alpha 3$ domain. The $\alpha 3$ and β_2 microglobulin associate together in a manner similar to that found in antibody structures. The function of the β_2 microglobulin is to stabilise the Class I molecules and it may also be involved in transporting heavy chains to the membrane.

In both Class I and Class II molecules, the antigen binding site is a groove approximately 20 Å long, 12 Å wide and 8 Å deep consisting of an eight-stranded β-pleated sheet topped by two antiparallel helices (Figure 1.5). In the Class I molecule the peptide binding site is formed by the N-terminal strands of α_1 and α_2 domains (Bjorkman *et al.*, 1987). It holds peptides in roughly extended conformation of about 8-10 residues with two or three conserved positions termed anchor residues (Romero, 1991; Falk, 1991; Hunt, 1992*a*; Jardetsky, 1991; Rotzschke, 1990; Van Bleek, 1990) and has allele specific binding motifs. Longer peptides can also fit by bulging partly out of the groove or by zigzagging within the cleft. In some cases the ends of the peptide may extend out of the groove although this decreases the stability of interaction.

Class I MHC molecules are extremely polymorphic particularly in the $\alpha 1$ and $\alpha 2$ domains and the most highly variable residues are found to point into this groove and up from the tops of both helices. These polymorphic residues are responsible for the specificity of recognition by T cells and for the variation in responsiveness to particular foreign antigens. Peptide sequence specificity is provided by six allele specific pockets A to F that bind side chains from the bound peptide.

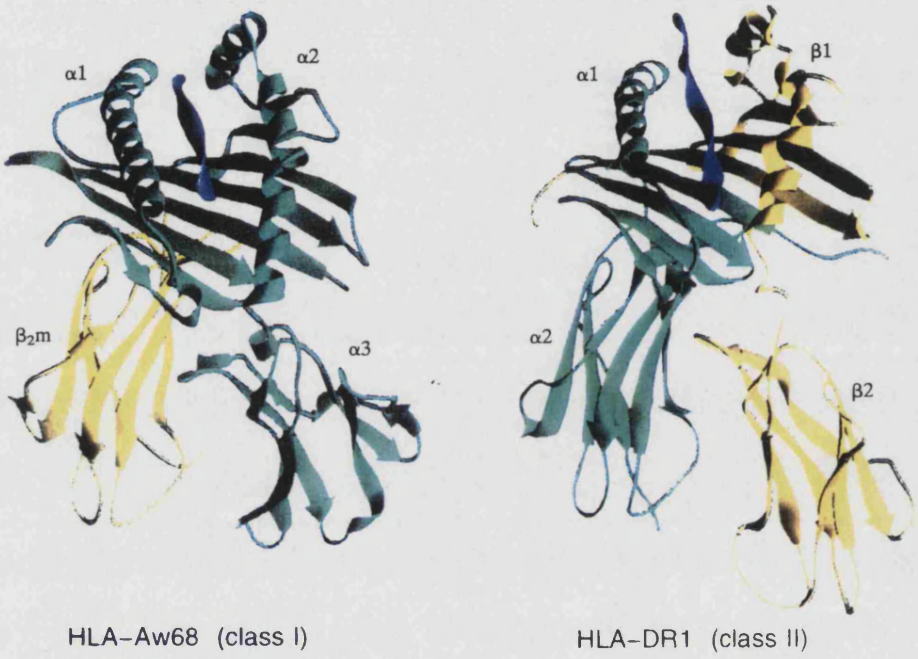


Figure 1.5 Comparison of MHC Class I and Class II crystal structures. Ribbon diagrams of Class I HLA-Aw68 and Class II HLA-DR1 proteins.The α (green) and the β domains (yellow) are shown. Peptides (blue) bind in the groove formed by the β-sheet and the two α helical regions. Although the domain organisation of Class I (α1α2α3, β_2m) and Class II (α1α2, β1β2) differ, their three dimensional structures are very similar. β_2m = $\beta_2$microglobulin.The figure is taken from Stern and Wiley (1994).

1.5.2.2 Class II MHC- peptide interactions

Class II molecules are expressed only on a subset of cells with specific immunological functions namely dendritic cells, macrophages, B cells and thymic epithelial cells. Expression of Class II molecules can be induced on other cell types by cytokines, interferon-γ and TNF, which also increases the expression of Class I molecules.

Class II molecules are heterodimers comprising heavy (α) and light (β) glycoprotein chains. Each chain consists of two extracellular domains (α1 and α2 or β1 and β2), a transmembrane region and a cytoplasmic domain. The membrane proximal domains α2 and β2 are structurally similar to Ig constant regions. The membrane distal domains α1 and β1 associate to form the peptide binding region. This region consists of eight strands of anti- parallel β-sheet with two anti-parallel α-helical regions overlaying them analogous to that seen between Class I α1 and α2 domains. There is a deep cleft between the α-helices that accommodate the bound peptide (Figure 1.5).

Despite their overall similarity, differences are seen in the antigen binding site of the Class II MHC molecule, HLA-DR1 and Class I MHC molecules. The grooves of Class I MHC molecules are blocked at both ends, while those of Class II molecules are open. This allows the Class II molecules to bind longer (13-25 residue) peptides (Rudensky *et al.*, 1991; Chicz *et al.*, 1992; Hunt *et al.*, 1992*b*; Newcomb *et al.*, 1993). In the Class I molecules, the groove tapers at both ends to a width of approximately 5 Å and is then blocked completely by bulky amino acid side chains, including Tyr 84 and Trp 167 that are conserved in virtually all Class I MHC

sequences. In the Class II molecule, HLA-DR1, the blocking side chains are replaced by smaller ones and are repositioned by secondary structural changes concentrated at the ends of the cleft. The conserved residues that bind the peptide N and C terminus in the Class I structure are also not found in Class II sequences. In the structure of HLA-DR1 complexed with a 13 residue influenza virus peptide, the peptide antigen binds as a straight extended chain with a pronounced twist (Stern *et al.*, 1994). Side chains of the peptide are accommodated by polymorphic pockets of the MHC groove. Unlike Class I, the N and C termini extend out of the groove and protrude into the solvent. However the area of peptide available for direct contact with TCR (400-500 $Å^2$) is the same for the majority of Class I -peptide complexes. However in the case of some Class I molecules this area is limited to 100-300 $Å^2$ (Fremont *et al.*, 1992).

1.5.3 T cell recognition of processed antigens

Antibodies recognise conformational epitopes exposed in the three dimensional structure of the antigen while T cells recognise sequential epitopes that are revealed when the molecule is unfolded or degraded.

The phenomenon whereby native antigen is converted to an altered or processed form that can be recognised by T cells is known as antigen processing. Class I molecules present peptides derived from endogenously synthesised antigens to $CD8^+$ T cells. These peptides are translocated in an ATP-dependent fashion into the lumen of the ER for binding to Class I molecules by means of the MHC encoded transporters associated with antigen processing TAP1 and TAP2 (members of a

superfamily of proteins containing an ATP-binding cassette) and form heterodimers in the ER. Two additional genes LMP2 and LMP7 which map close to TAP1 and TAP2 encode products that are related to subunits of a large cytoplasmic complex called the proteasome known as prosome, macropain or multi-catalytic protease. The proteasome is involved in the degradation of proteins into small peptides which are then transported into the ER (Figure 1.6).

Class II molecules show a converse pattern, being inhibited from binding peptides within the ER and being specialised to present peptide antigens derived predominantly from extracellular sources and degraded within intracellular vesicles (endosomes or lysosomes).

They associate with fragments of proteins in the endocytic route and present them to $CD4^+$ T cells. Class II α and β chains associate with a third subunit, the invariant chain (Ii) [Kvist *et al.*, 1982; Machamer and Cresswell, 1982; Cresswell *et al.*, 1987] early during biosynthesis. This heterotrimeric complex is transported from the ER to the trans-Golgi reticulum where it is sorted to endocytic compartments (Cresswell *et al.*, 1985; Neefjes *et al.*, 1990; Lamb *et al.*, 1991). Association of Class II molecules with Ii inhibits the binding of peptides to the Class II molecule, thus preventing binding of peptides within the ER, while limited proteolysis of the invariant chain within the endosomal compartment releases the inhibition and allows peptide binding. Proteolysis of Ii frees the Class II molecule to be transported from the endosomal compartment to the cell surface. The Class II molecule is prevented from binding peptides transported into the ER and is targetted to a compartment where proteolysis of endocytosed extracellular proteins occur.

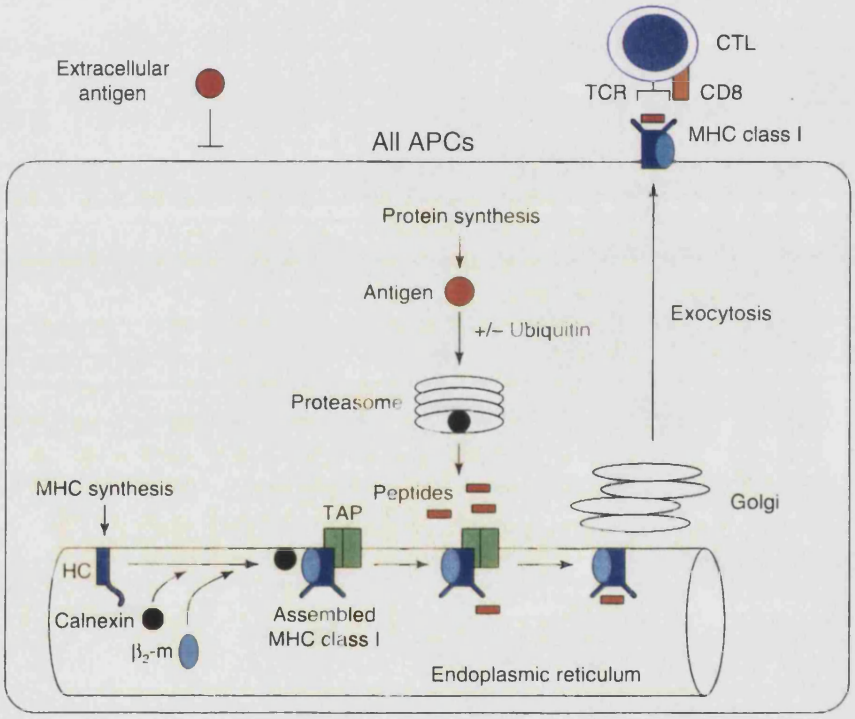


Figure 1.6 The classical pathway for presentation of endogenous antigen on major histocompatibility complex (MHC) class I by antigen-presenting cells (APCs) (taken from Rock *et al.*, 1996).

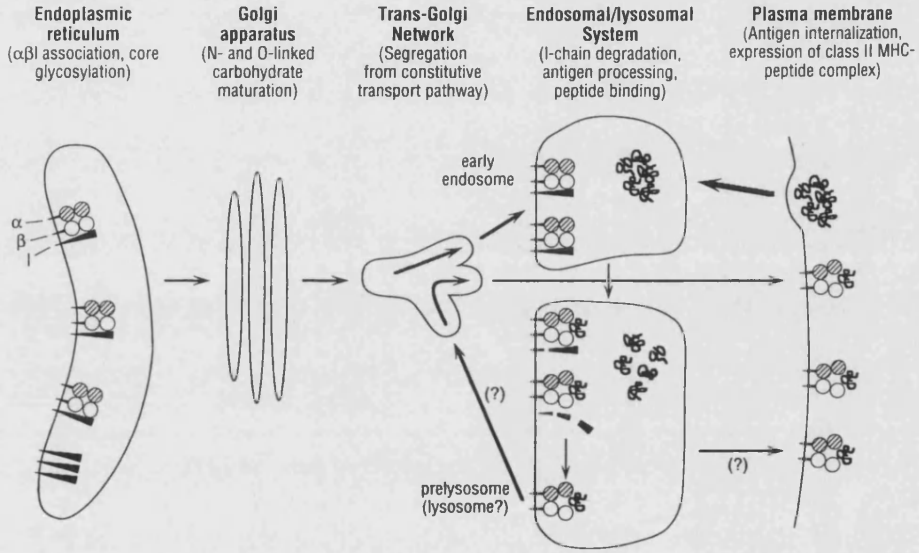


Figure 1.7 Schematic view of the assembly, intracellular transport and peptide loading of MHC Class II molecules (taken from Cresswell., 1994).

1.5.3.1 TCR - peptide interactions

The interaction of the TCR with its ligand, a peptide in combination with a molecule of the MHC is a central event in the initiation and propagation of most immune responses to protein antigens. Specific αβ heterodimers present on the surface of most T cells are responsible for the recognition of antigenic peptides bound to Class I and II molecules (Davis *et al.*, 1988; Jorgensen *et al.*, 1992). It is clear from the crystal structures of Class I and II molecules that the polymorphic residues of MHC molecules are focused on interactions with peptide and not directly involved in TCR interactions.

A model was proposed by Davis and Bjorkman (1988) wherein the V(D)J junctional region of the TCR (equivalent to the CDR3 region of Ig) is principally responsible for peptide contact whereas CDR1 and CDR2 loops contact the α-helices of the MHC. This hypothesis explains the diversity in the CDR3 regions of the TCR and is compatible with the three dimensional studies of the MHC molecules and TCR models (Chothia *et al.*, 1988; Claverie *et al.*, 1989). Consistent with the model several studies have shown that the bound peptide contacts the CDR3 regions of both α and β chains and that these interactions are highly specific.

The recent structure of the β chain of the murine TCR (Bentley *et al.*, 1995) specific for a hemagglutinin peptide of influenza virus presented by the Class II molecule I-E^d showed restricted conformation heterogeneity in CDR1 and CDR2, which is consistent with the model of peptide-MHC recognition. However CDR1 and CDR2 were found to be closely associated with CDR3 which implies that the

three CDRs may not function independently. Obviously crystallographic studies of more TCR/peptide/MHC complexes will shed light on these interactions.

A strategy for investigating peptide-TCR interactions was developed by Jorgensen *et al* (1992). Analogs of moth cytochrome c (MCC 88-103) peptides altered at residues likely to contact the TCR were used to select for compensating changes in TCR structure. Many hybridomas specific for these altered peptides were generated and the analysis of their TCR sequence has given additional information on the nature of specificity. An important constraint that is employed in the system is the use of mice transgenic for either the α or β chains of a cytochrome c/ I-E^k reactive TCR (5C.C7). Either the α or the β chain TCR transgenic responds best to a substitution at a given position but not both. Peptide reactivities can be correlated with changes within the endogenous (non transgenic) TCR chain. T cells generated from normal mice which are specific for MCC peptide (88-103) bound to I-E^k use Vβ3 and Vα11.1 almost exclusively (Hedrick *et al.*, 1988). Although a different Jα and Jβ are used there is always a conserved acidic residue glutamate or aspartate in the CDR3 region of the α chain and a conserved asparagine in the CDR3 region of the β chain. Lysine which is at position 99 in the cytochrome c peptide most likely contacts the TCR. Peptides with a K/E change at this position elicit responses from T cells that utilise the transgenic B chain (Vβ3) together with Vα11.1 or Vα11.2 with the conserved D/E replaced by a K residue. A charge-charge interaction or 'salt bridge' between the E residue in the Vα CDR3 and the K on the peptide can be reversed in the case of 99E variants. T cell hybridomas that recognise E residue at position 99 of the peptide can also recognise D at this position but none of the other

18 possible amino acid substitutions. These results indicate that there is a direct CDR3-peptide interaction such that changes in a T cell determinant of the MCC peptide always alters the responding TCR Vα or Vβ CDR3 sequence. Peptide-TCR and TCR- MHC contacts are highly dependent on each other.

1.5.3.2 TCR-MHC interactions

Ehrich (1993) used structurally very similar and well characterised panels of T cells generated by Jorgensen *et al* (1992) to survey T cell responses to peptide analogs presented by mutant I-E^k molecules. Each I-E^k mutant consists of a single amino acid substitution at a residue predicted to point up from the α-helical region of either the E_α or the E_β chains. The system was designed to evaluate the effect on the overall MHC-TCR interaction by small defined changes in either the TCR-CDR3 region or in the peptide. Cross reactive T cell hybridomas show very different sensitivities to many of the I-E_α^k and I-E_β^k mutant depending on which of the two peptides is being recognised. The same TCR can contact the same MHC in different ways. Very distant TCR-MHC contacts can be affected by changes in CDR3-peptide contacts. TCR-MHC contacts are 'fluid' which means that they can be made in a variety of ways with the same TCR and MHC. The final configuration appears to be significantly influenced by the peptide. The TCR-peptide interactions dramatically change the overall TCR-MHC interactions. TCR mediated recognition of a peptide-MHC complex is 'peptide centred' and CDR3-peptide interactions determine the ultimate configuration.

1.6 Structure/ function analysis of T cell specificity

That a single TCR αβ heterodimer subserves the complete recognition of both nominal antigen and MHC molecules in MHC-restricted responses has been demonstrated conclusively for both Class I and Class II restricted recognition by gene transfection studies. In T cell clones with dual specificity for antigen/self MHC and allogeneic MHC, a single TCR αβ heterodimer has been shown to mediate both the MHC- restricted antigen recognition and the allorecognition (Malissen *et al.*, 1988; Kaye *et al.*, 1988). A common approach to exploring the structural basis of T-cell recognition has been to correlate TCR usage with antigen and MHC specificity in well characterised immune responses. Many studies have demonstrated such correlations to varying degrees for both Vα and Vβ expression. Predominant expression of murine Vα 11 family genes has been observed in the I-E subregion-restricted response to the antigen pigeon cytochrome c (Fink *et al.*, 1986; Winoto *et al.*, 1986), of Vα3 gene segments in the response to the hapten p-azobenzenearsonate (Tan *et al.*, 1988), of Vα3 in the I-E^k restricted response to the λ repressor fragment 12-26 (Lai *et al.*, 1988), of Vα4 in the TNP/ H-$2K^b$ restricted cytotoxic T cell clones and Vα4 and Vα2.3 among murine T cell clones specific for the encephalitogenic fragments p1-9 and p1-11 of MBP in PL and B10.PL mice, respectively. Similarly Vβ elements have also been found to be preferentially used. Predominant expression of Vβ8 in the I-E^d restricted response to sperm whale myoglobin (Morel *et al.*, 1987), Vβ3 in the H-$2K^b$ /TNP response, Vβ 8.2 on encephalitogenic T cell clones in the response to mbp fragments and Vβ3 in the

cytochrome c response (Fink *et al.*, 1986; Winoto *et al.*, 1986). Analysis of Vβ usage among many T cell clones that are I-E-restricted and pigeon cytochrome c has revealed a striking correlation between V_β expression and the pattern of clonal MHC restriction. These data suggested that there may be duality of function of TCR α and β proteins such that the Vβ has a prominent role in MHC recognition while Vα mediates specificity for the nominal antigen. However subsequent analyses have shown that TCR mediated recognition cannot be explained simply by correlating Vβ expression with MHC specificity and Vα expression with antigen specificity. Examination of TCR usage in antigen specific responses can be extended to responses to alloantigens.

1.6.1 Allorecognition

1.6.1.1 Theories to explain allorecognition

In contrast to the low frequency of T cells that recognise conventional antigens, 1-10% T cells respond to an alloantigen. One of the earliest attempts to explain the basis for the high frequency of alloreactive cells was described by Jerne (1971). The repertoire of lymphocyte receptors was evolutionarily preselected to include germline genes that had specificity for the MHC molecules of the species. According to the Jerne theory, for each MHC molecule there evolved a complementary receptor molecule and the receptor repertoire used in recognition of foreign antigen was derived by somatic mutation of the receptor specific for self MHC molecules. This avoids autoreactivity and provides the diverse repertoire required for immune surveillance. Another hypothesis was put forward by

Matzinger and Bevan (1977) to explain the high frequency of alloreactive cells. A single type of allogeneic MHC molecule can give rise to multiple distinct specificities by forming binary complexes with an array of peptides derived by internalisation and processing of endogenous molecules. Each of these peptide-MHC complexes could then stimulate a separate population of T cells and hence account for the high precursor frequency of alloreactive cells. No restriction was placed on the extent of diversity permitted to the TCR repertoire (Matzinger, 1977). Initial attempts to assess diversity were carried out in a series of mutant mice that differed from wild type (C57BL/6) in the K^b molecule. The TCR repertoire to a small alloantigenic difference was complex and estimated to comprise at least 37 different TCRs . One or several Vβs recur with high frequency within a particular allo-response. Allospecific CTL were part of the repertoire devoted to recognition of antigen in the context of self MHC. Clones restricted by self MHC were found to display alloreactivity in some cases (Matis *et al.*, 1987; Sredni *et al.*, 1980; Braciale *et al.*, 1981). Molecular analysis demonstrated that the same portion of the TCR involved in antigen recognition was involved in recognition of alloantigen.

Another hypothesis, the high determinant hypothesis, was put forward by Bevan (1984). The other theories were based on the assumption that the epitopes recognised by alloreactive cells are equivalent to those seen in conventional T cell responses and the difference in the magnitude of these responses was due to the large number of epitopes present on an alloantigen. Conventional foreign antigens must compete with the multitude of cellular epitopes to interact with MHC for presentation to T cells and therefore would be found associated with relatively few

Class I molecules While in the case of an alloantigen, the foreign sequence on an allo MHC is present on all Class I molecules and is therefore abundant. The high concentration of antigen could permit stimulation of many T cells of relatively low affinity. Thus the difference between allo T cells and those responsive to conventional antigen was related to the range of affinity of the receptors that were stimulated. Receptors reactive to conventional antigens should have a high affinity which is seen in only a few T cells while alloreactive T cells would require a lower affinity which is seen in a higher proportion of the receptor repertoire.

1.6.1.2 Contribution of MHC bound peptides to allo recognition

Peptide involvement in allorecognition was first observed in Class I molecules (Heath *et al.*, 1989). A panel of K^b specific CTL clones was tested for the ability to lyse a human cell line transfected with K^b (Jurkat K^b). This analysis sought evidence for peptide dependence by searching for a CTL clone that recognised a murine peptide not expressed in human cells. Although the majority of K^b specific clones were able to recognise K^b expressed by the human target several could not. To check if this lack of recognition was due to lack of presentation of a specific cellular peptide, cytoplasmic proteins were prepared from a murine cell line cleaved with cyanogen bromide to produce peptides. This demonstrated that at least some T cells were peptide dependent in their recognition of alloantigen. Results of Townsend *et al* (1989) provide strong support for this suggestion. Working with a mutant cell line RMA-S which is almost devoid of cell surface MHC Class I expression they concluded that $\alpha\beta$ assembly, transport to the cell surface and the

native conformation of Class I molecules are dependent on bound peptide. Pulsing the mutant RMA-S cells with high concentration of an influenza NP peptide known to bind H-2D^b restored cell surface expression of D^b to 20% of normal levels. NP occupied Class I molecules appear to be conformationally normal. Peptide involvement was also seen in the case of Class II allorecognition. Incubation of Class II presenting cells with exogenous antigen was observed to reduce the ability of alloreactive T cells to recognise Class II MHC molecules. High concentration of exogenous antigen inhibited presentation of a different peptide required for allorecognition.

Peptide mediated inhibition or enhancement of allorecognition has also been observed. Parham *et al.*, (1987) described inhibition of human anti HLA -A2 cytotoxic T cells by addition of peptides comprising sequences from N-terminal domains of HLA-A2. It is not clear whether this effect results from occupancy of a site on the TCR or from displacement of a peptide from the antigen binding groove of HLA-A2. The inhibitory or enhancing effects of an influenza haemagglutinin peptide which is known to bind to DR1 on the responses of anti-DR1 human T cell clones has been studied by Eckels *et al* (1988). Either the displacement or mimicking of co-recognised endogenous peptides could lead to positive or negative effects respectively.

Site directed mutagenesis has been used to study the role of specific residues in allorecognition. HLA-DRB1*0103 differs from HLA-DRB1*0101 by 6 base pairs clustered in the third variable region of the second exon leading to three amino acid changes at positions 67, 70 and 71 of the beta chain of the HLA-DR

molecule. SDM was used to examine the role of each residue. Residue 71 in the alpha helix of the DR beta domain was found to influence peptide binding and plays a central role in allorecognition. It was also observed in an influenza HA specific HLA-DRB1*0103 restricted T cell line (Coppin *et al.*, 1993).

Generation of an allodeterminant by the addition of a peptide has also been seen. Cells expressing HLA-Aw69 were converted into targets for an HLA-A2 specific alloreactive T cell by the addition of a peptide containing sequence from the region of the α- helix of the α_1 domain of A2 that differs from Aw69 (Clayberger *et al.*, 1987). All these observations suggest that a single foreign MHC molecule can give rise to multiple binary complexes and thus stimulate multiple alloreactive T cells, each with a unique specificity.

1.6.1.3 The diversity of TCR genes used in allorecognition

TCR Vα and Vβ gene usage has been studied by a number of groups. Non-random patterns of TCR selection have been seen in a number of cases. The molecular diversity of the allogeneic TCR repertoire specific for the HLA-DR4 variant DRB1*0404 was studied by Goronzy *et al* (1993). A core group of Vβ elements (Vβ 6, 5, 2, 13.2, 18 and 7) was found to be preferentially used. Another study of Vβ selectivity in HLA-Dw14 alloreactive T cell clones showed the usage of Vβ 8.2. TCR genes from clones derived *in vitro* from mixed lymphocyte cultures between Dw4+ and responder and Dw14+ stimulator cells also showed non-random gene usage (Yamanaka *et al.*, 1993). In the case of DR1 microvariation non-random TCR usage has been seen. Six different Vα and seven Vβ segments are utilised by 9

different DR (α,β1*0101) specific TLC which suggests that the TLC do not appear to require specific V segments to recognise a microvariant of the DR1 alloantigen. However since the same Vα and Vβ segments are found in more than one TLC and these TLC are derived from different individuals, the usage of V segments in these TCR appears to be non-random. (Hurley *et al.*, 1993). However since a majority of the amino acid sequence and the 3-D structure is shared by the two microvariants DR (α,β1*0101) and DR (α,β1*0102) the use of the same V segments is not unexpected. Five of these TCR sequences have been modelled in this thesis and they are discussed in more detail in Chapter 3. A CD4+ T cell clone that recognises HA 306-320 in the context of autologous DR1101 molecules as well as of allogeneic DR1301, DR0402, DR1501 and DR1601 molecules was studied by (Zeliszewski *et al.*, 1993). Degenerate recognition was mediated by a single TCR as judged by both V- gene sequencing and competition assays. Substitutions of DR residues within the third HYR region resulted in a loss of T cell reactivity which is restored by additional substitutions in the first and second HYRs. Heterogeneity was also observed in the case of murine I-Abm 12 reactive T cells (Bill *et al.*, 1989). In this case the repertoire to the alloantigen was estimated to be at least 37 different TCRs.

1.7 This Thesis

Initial aims

One of the original aims of this study was to determine and then analyse the sequences and 3-D models of TCR Vα and Vβ genes used in peptide specific DR4 Dw4 restricted clones. A strategy to sequence TCR Vα and Vβ genes from DR4

Dw4 restricted clones was devised. This strategy involved the amplification of the Vα and Vβ genes by a modification of PCR known as Anchor-PCR. Conventional PCR could not be used for the amplification of the TCR Vα and Vβ genes as the 5' portion of the sequence is always unknown. Anchor PCR requires sequence specificity only on the 3' end of the target fragment (Figure A2.3).

Typically, total RNA is isolated from a T cell clone and the first strand of cDNA synthesized using oligo $(dT)_{12-18}$ primer. A poly dG tail is then introduced at the 5' end of the strand with an enzyme terminal deoxynucleotidyl transferase. The product is then amplified with a specific Cβ 3' primer and another oligonucleotide consisting of a poly dC tail attached to a sequence with a Not I restriction site termed the 'anchor'. The amplified products are then cloned into M13mp18 and sequenced by Sanger's dideoxy chain termination method.

Another mechanism which bypasses the cloning steps was also devised. This method called PCR Sequencing allows the direct sequencing of amplified products. The Anchor-PCR amplified products were screened by Southern blotting to ensure their validity. These Southern blotted products were then subjected to PCR Sequencing. This method is a variation of Sanger's dideoxy method. It involves an annealing step where the template-primer is snap-frozen immediately after boiling the mixture. Subsequent labelling and termination reactions are similar to Sanger's method.

This method was optimised in the T cell line KM vs Priess. However it could not be extended to peptide specific DR4 Dw4 restricted clones, supplied from another laboratory, as they were found to be contaminated with mycoplasma late in

the development of the sequencing work. Contamination of cell cultures by mycoplasma caused degenerative effects on TCR mRNA as well as ubiquitously expressed mRNAs such as glyceraldehyde 3-phosphate dehydrogenase (G3PDH). Control reactions using cross species G3PDH specific primers were used to detect this contamination.

Revised aims

Subsequently the aims of this project were revised to analyse the TCR Vα and Vβ gene sequences from DR restricted clones found in the sequence databases and the literature to generate 3-D models of the TCRs using a modified version of the antibody modelling software, AbM, namely TCRM (Searle and Rees, unpublished). The purpose of this analysis would be to attempt to establish a relationship between MHC restriction and TCR sequence and structure.

Chapter 2 describes the sequence analysis of TCR Vα and Vβ genes from Class I and II restricted clones found in the database. These sequences were analysed using a number of Genetic Computer Group (GCG) programs in an effort to understand TCR usage in the various clones. The CDRs of TCR Vα and Vβ genes from Class I and II restricted clones were analysed and conserved motifs or residues were identified.

Chapter 3 presents the molecular models of TCR VαVβ heterodimers generated from DR restricted clones by TCRM.

Chapter 4 discusses all the findings in the study of the TCR VαVβ heterodimer in the context of present hypotheses.

At the beginning of the PhD. studies several projects were looked at to gain experience in methods of molecular biology. The two major pieces of work begun are described in Appendices 1 and 2.

Appendix 1 describes the expression and purification of a catalytic antibody double aspartate mutant Tyr32AspLeu92Asp. This antibody was expressed in *E.coli* as Fv fragments and purified by affinity chromatography by a modification of the method used by Hilyard, 1991.

Appendix 2 describes the materials and methods used in the cloning and sequencing of TCR genes from a T cell line KM vs Priess.

Time table

October 1990 - June 1991	-	Gloop2 Fv (catalytic mutant) expression and purification.
June 1991 - September 1994	-	KM vs Priess TCR cloning.
September 1994 - April 1995	-	Short break to India while awaiting preparation of T cell clones.
April 1995 - August 1996	-	TCR Sequence Analysis and Modelling.

CHAPTER 2

Sequence Analysis Of TCR Variable Domains

The comparative analysis of protein sequences is the first step in the study of protein structure and function. When this is coupled to three-dimensional information for a given family of homologous proteins it becomes a powerful tool for determining residues which are important for a particular structural or functional role.

Although the complete three-dimensional structure of the TCR VαVβ heterodimer has not yet been determined, the separate Vα and Vβ structures are known along with a large number of TCR Vα and Vβ gene sequences.

2.1 Acquisition of TCR Vα and Vβ sequences

A specific TCR sequence database was constructed using data from available DNA and protein databanks (EMBL, NEWAT, NBRF-PIR and KABAT) of aligned human Vα and Vβ gene sequences. The sequence alignments were performed on the alpha and beta chains separately depending on the MHC restriction and peptide specificity. The sequences were obtained via anonymous ftp and converted to Genetics Computer Group (GCG) format for further analysis.

2.2 *Variability analysis*

Variability was calculated for each chain type for sequences which contained V, D and J segments using the method of Wu and Kabat (1970). The overall variability at each alignment position is calculated as the number of residue types occurring at the position divided by the frequency of the most commonly occurring residue at that position.

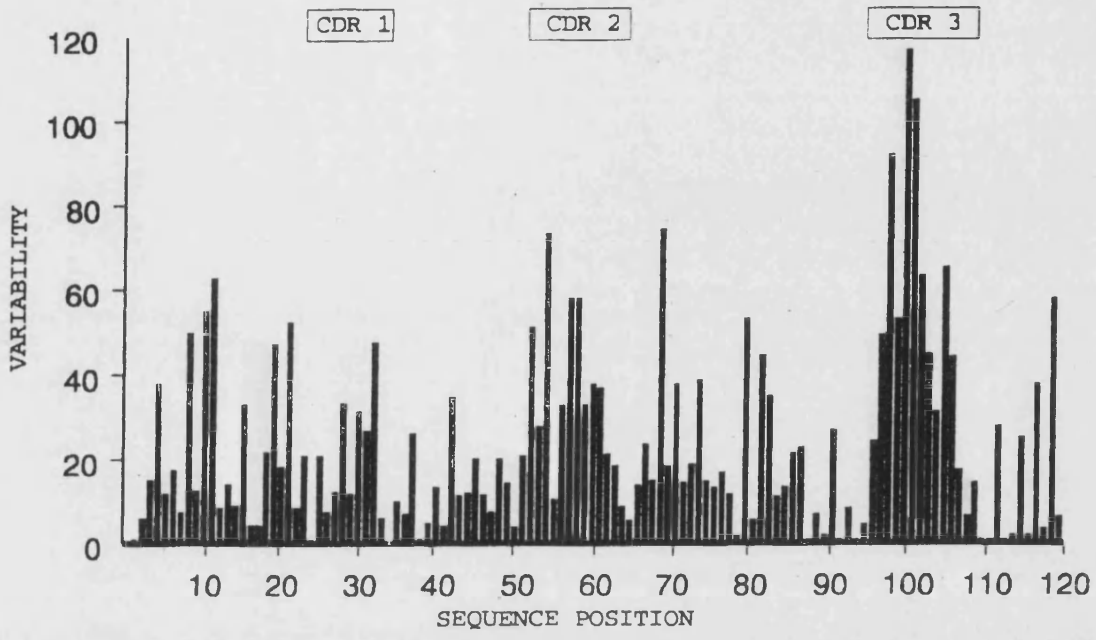
Variability plots show the high overall variability present in the TCR Vα and Vβ sequences (Figure 2.1). Recognisable peaks are present in the areas equivalent to second and third CDRs of antibodies and to a lesser extent in the first CDR region. A further region of high variability within a loop (the D-E loop) between the CDR2 and CDR3 equivalent regions of both chains is present. This suggests that this D-E loop region has less conserved structure than in antibodies and it may therefore form part of the TCR combining site.

2.3 *Sequence alignment*

The first step in any modelling is to convert the DNA sequence available into the primary structure of the gene product and search for sequence homologies with known sequences of other proteins. Information on DNA sequences is available in databanks such as those of the European Molecular Biology Laboratory (EMBL), Heidelberg, National Biomedical Research Foundation, Maryland or the NEWAT databank in the USA.

Sequence databases facilitate comparisons of newly determined sequences as well as act as a source of data for the generation and testing hypotheses about

α chains



β chains

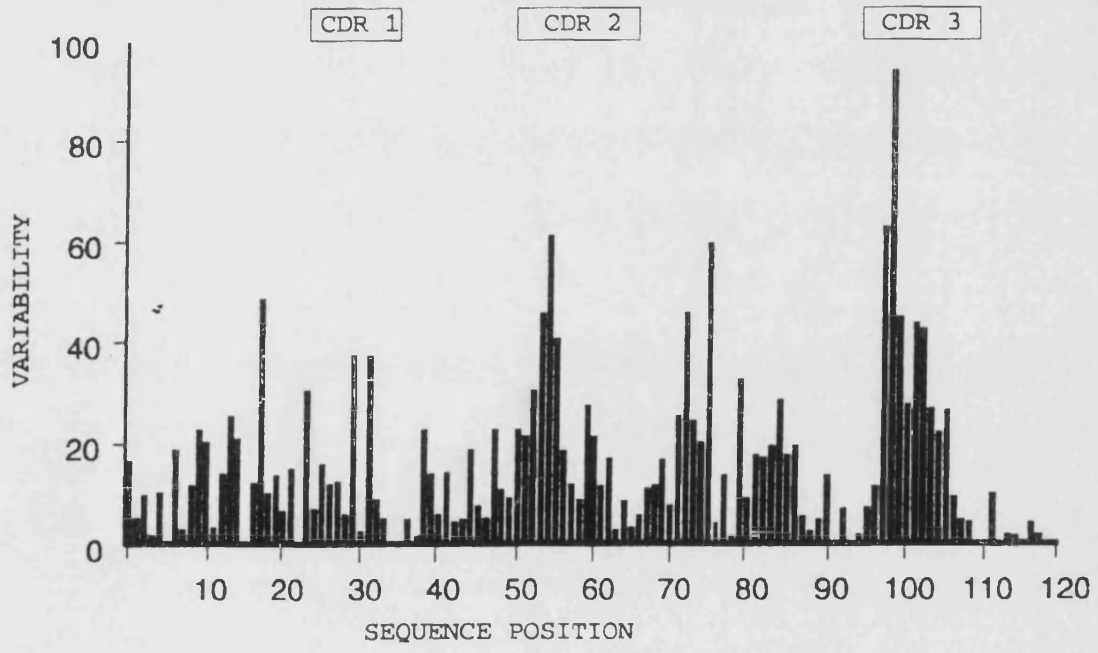


Figure 2.1 Variability plots of the TCR Vα and Vβ sequences. Apart from the three CDRs, a fourth variable region is seen in between CDR2 and CDR3 in both cases (the D-E loop).

molecular sequence organisation, function and evolution.

The major protein sequence databases are NBRF-PIR (National Biological Research Foundation- Protein Identification Resource) database and the PSD-Kyoto (Protein Sequence Database-Kyoto) database.

The PIR originated from the work of Margaret Dayhoff and was established at the National Biological Research Foundation in 1984. The NEWAT database was compiled before the PIR was available in order to examine protein relationships. It is a useful subset for rapid retrieval. There are also a number of specialized collections such as the KABAT database of sequences of proteins of immunological interest published by the US National Institute of Health and the International Haemoglobin Information Centre. PGtrans is a computer-generated collection of protein sequences obtained by the translation of GenBank. PseqIP is a non-redundant collection derived from PIR, PSD, NEWAT and PGtrans.

The three major DNA sequence databases include the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory Nucleotide Sequence Data Library (EMBL) and the GenBank Genetic Sequence Data Bank (GenBank).

The National Biological Research Foundation provides some software to search their databanks, to compare user-specific segments with segments of the same length in the database, to align two sequences, to detect similarities and to score and display the degree of similarity.

When homology is found between more than two proteins or nucleotide sequences, additional information can be obtained if a multiple alignment of all sequences is obtained. Accurate multiple sequence alignment highlights the residues

of common functional and structural importance. The location of identities and conservative substitutions can be used to guide the design of site-directed mutagenesis experiments whilst the identification of subtle patterns of residue conservation can yield improvements in the accuracy of secondary and tertiary structure predictions. Identification of conserved residues can suggest a local region that contains a motif that specifies a particular structure or function.

Programs used in this study include the Wisconsin Sequence Analysis Package by Genetics Computer Group (GCG) Inc. It is an integrated package of over 130 programs that allows nucleic acid and protein sequences to be manipulated and analysed. This was originally built by John Devereux and Paul Haeberli. Each program is designed to perform one function and complex analyses are performed by using several programs in succession. Tasks can be performed by selecting specific tools. The programs are loosely divided into six groups namely sequence comparisons, restriction enzyme and chemical and proteolytic cleavage site mapping and searching, prediction of nucleic acid secondary structure and protein secondary structure, pattern recognition and sequence manipulation.

A number of multiple sequence alignment GCG programs including PILEUP, LINEUP and PRETTY were used in the study of TCR Vα and Vβ sequences with different MHC and peptide specificities. These programs are described below.

2.3.1 PILEUP

The PILEUP program from the GCG package employs a simplified version

of the progressive pairwise alignment method of Feng and Doolittle (1987) and performs a multiple sequence alignment from a group of related sequences creating a multiple sequence format (MSF) file. This method is similar to the method described by Higgins and Sharp (1989).

The alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster can then be aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences can be aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments that include increasingly dissimilar sequences and clusters until all sequences have been included in the final pairwise alignment.

Before alignment, the sequences are first clustered by similarity to produce a dendrogram or a tree representation of clustering relationships. This dendrogram directs the order of the subsequent pairwise alignments. The clustering strategy represented by the dendrogram is called UPGMA (unweighted pair-group method using arithmetic averages); Sneath and Sokal, 1973. This dendrogram is not a phylogenetic reconstruction although the vertical branch lengths are proportional to the distance between the sequences. The distance along the horizontal axis has no significance.

It can align up to 500 sequences with any single sequence in the final alignment restricted to a maximum length of 7000 characters including gap characters inserted into the sequence by PILEUP to create the alignment. If longer

sequences are included in the alignment, the number of sequences that PILEUP can align are decreased. It is an extension of the Gap program for more than two sequences. It tries to find global optimal alignment rather than a local optimal alignment. It can take more than a few minutes to run depending upon the length and number of sequences being aligned and so it is preferable to run them in batch.

PILEUP was used to perform a multiple sequence alignment of the TCR Vα and Vβ sequences of all the possible combinations of peptide and MHC restriction.

The output file created was fed into the PRETTY program which displays a consensus sequence for the aligned sequences on the basis of threshold and plurality values which define the number of identical residues which have to exist at a particular position for a consensus to be valid.

2.3.2 LINEUP

This program edits multiple sequence alignments and about 30 sequences can be edited simultaneously. A consensus sequence identifies places where the sequences are in conflict.

A number of computer programs have been developed to aid the interpretation of multiple sequence alignments. GCG programs used for the analysis of TCR sequences include PRETTY, PLOTSIMILARITY, FIGURE and SEQED. PRETTY displays multiple sequence alignments and calculates a consensus sequence for the sequence group created by LINEUP. It does not create the alignment but simply displays it. It can only handle sequence groups whose left ends

are aligned. PLOTSIMILARITY plots the running average of the similarity among the sequences in a multiple sequence format. FIGURE creates figures by including both graphics and text. SEQED facilitates the entry and modification of sequences as well as assembly of parts of existing sequences into new genetic constructs.

2.4 Sequence comparison

The large number of TCR Vα and Vβ sequences in the database can be roughly divided into four groups depending on their MHC restriction and peptide specificity. The four groups are outlined in Table 2.1. The PILEUP analyses of these sequences are shown in Figures 2.2-2.11.

The diversity of the sequences in these four groups was analysed by calculating the APC and MPI values.

2.4.1 MPI and APC analysis

A group of sequences can be characterised by a 'minimum pairwise identity' (MPI) value which is the percentage identity shared by the two proteins in the group that are the least similar (Benner, 1989). TCR Vα and Vβ sequences in the four groups were further separated into sets depending on the peptide and MHC restriction. These sets were then aligned using PILEUP (Figures 2.2-2.11). Since the sequences are clustered according to similarity in PILEUP, the two sequences at the opposite ends in a multiple sequence alignment are least similar. The percentage identity of the end sequences in each set was calculated to give the MPI value of that set.

Group	Specificity
Group 1	same peptide, same MHC
Group 2	same peptide, random MHC
Group 3	random peptide, allo MHC
Group 4	random peptide, random MHC

Table 2.1 Groups of TCR Vα and Vβ sequences

```
PileUp of: @mouse.flu.alpha

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 mouse.flu.alpha.pileup.msf  MSF: 18  Type: P  November 28, 1995
21:57  Check: 5821 ..

//

        1     CDR3       18
u12155  CAV.....YA NKMIFGLG
u12154  CALSEWGSSG NKLIFGIG
u12166  CAV..RGTNA YKVIFGKG
u12169  CAVGDTGANT GKLTFGHG
u12164  .CALG..DSN YQLIWGSG
u12161  CALMK..DSN YQLIWGSG
u12157  .CVLSARDSN YQLIWGSG
u12163  .CALGESGS. WQLIFGSG
u12159  ....CAANSN NRIFFGDG
u12156  CAVSMAGTRT S..ILGKG
```

Figure 2.2 PILEUP of murine TCR Vα sequences from flu restricted clones [Group 1]. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: @mouse.beta.lcv.pep

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 mouse.beta.lcv.pileup.msf  MSF: 14  Type: P  November 28, 1995
23:28  Check: 1673 ..


//

       1           14
m61140  .CASSSGGNY AEQF
m61139  .CASSYWGDY AEQF
m61138  CASSSTRDSY AEQF
m61141  CASSSGLGGY .EQY
```

Figure 2.3 PILEUP of murine TCR Vβ sequences from lcv restricted clones [Group 1]. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: mbp.alpha.pep.msf{*}

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                GapWeight: 3.000
           GapLengthWeight: 0.100

 mbp.alpha.pep.pileup.msf  MSF: 149  Type: P  November 28, 1995
22:20  Check: 5905 ..

 //

       1                                                          <
x58336  ......... ......... LLGRTRGDSV TQMEGPVTLS EEAFLTINCT
x57531  ......... ........LL LLGRTRGDSV TQMEGPVTLS EEAFLTINCT
x57534  .......MAS APISMLAMLF TLSGLRAQSV AQPEDQVNVA EGNPLTVKCT
x58334  .MAGIRALFM YLWLQLDWVS R.....GESV GLHLPTLSVQ EGDNSIINCA
x57533  .......SVL ILWLQPDWVN SQQKNDDQQV KQNSPSLSVQ EGRISILNCD
x61628  ......... ......... .......KEV TQIPAALSVP EGENLVLNCS
x57532  ......... ......... ........EV TQIPAALSVP EGENLVLNCS
x58333  METLLGVSLV ILWLQLARVN S......QQG EEDPQALSIQ EGENATMNCS
x58322  ......... MRLVARVTVF LTFGTIIDAK TTQPTSMDCA EGRAANLPCN
x61629  .......MSN QVLCCVVLCF LGANTVDGGI TQSPKYLFRK EGQNVTLSCE


          CDR1    >                < CDR2  >              100
x58336  Y.TATGYPSL FWYVQYPGEG LQLLLKATKA DD..KGSNKG FEATYRKETT
x57531  Y.TATGYPSL FWYVQYPGEG LQLLLKATKA DD..KGSNKG FEATYRKETT
x57534  Y.SVSGNPYL FWYVQYPNRG LQFLLKYITG DNLVKGSY.G FEAEFNKSQT
x58334  Y.SNSASDYF IWYKQESGKG PQFIIDIRSN MDKRQGQR.. VTVLLNKTVK
x57533  Y.TNSMFDYF LWYKKYPAEG PTFLISISSI KDKNEDGR.. FTVFLNKSAK
x61628  F.TDSAIYNL QWFRQDPGKG LTSLLLIQSS QREQTSGR.. LNASLDKSSG
x57532  F.TDSAIYNL QWFRQDPGKG LTSLLLIQSS QREQTSGR.. LNASLDKSSG
x58333  Y.KTS.INNL QWYRQNSGRG LVHLILIRSN EREKHSGR.. LRVTLDTSKK
x58322  HSTISGNEYV YWYRQIHSQG PQYIIHGLKN NETNEMA... .SLIITEDRK
x61629  ..QNLNHDAM YWYRQDPGQG LRLIYYSQIV NDFQKGDIAE GYSVSREKKE


       101                 <     CDR3       >           149
x58336  SFHLEKGSVQ VSDSAVYFCA .....LRG.G TDKLIFGTGT RLQVFPNIQ
x57531  SFHLEKGSVQ VSDSAVYFCA .....LIG.F GNVLHCGSGT QVIVLPHIQ
x57534  SFHLKKPSAL VSDSALYFCA .....VRDTG ANNLFFGTGT RLTVIPYIQ
x58334  HLSLQIAATQ PGDSAVYFCA E.....KGF. .QKLVFGTGT RLLVSPNIQ
x57533  HLSLHIVPSQ PGDSAVYFCA A.....SATF GNKLVFGAGT ILRVKSYIQ
x61628  RSTLYIAASQ PGDSATYLCA V...PLQAGA NSKLTFGKGI TLSVRPDIQ
x57532  RSTLYIAASQ PGDSATYLCV C......SGT ASKLTFGTGT RLQVTLDI.
x58333  SSSLLITASR AADTASYFCA P...L...GG GNKLTFGGGT RVLVKPNIQ
x58322  SSTLILPHAT LRDTAVYYCI V...RAYSG. NTPLVFGKGT RLSVIANIQ
x61629  SFPLTVTSAQ KNPTAFYLCA SSIRVFWDSR NTEAFFGQGT RLTVVED..
```

Figure 2.4 PILEUP of myelin basic protein specific human TCR Vα sequences (without sequenced Vβ partners) [Group 2]. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: @mbp.beta.pep

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 mbp.beta.pileup.msf  MSF: 135  Type: P  November 28, 1995 22:39
Check: 2737 ..

//

        1                                                  <   CDR1
x58325  .......... FCLLQAGPLD TAVSQTPKYL VTQMGNDKSI KCEQNLGHDT
x57535  ..CRLLCCVV FCLLQAGPLD TAVSQTPKYL VTQMGNDKSI KCEQNLGHDT
x58324  MDSWTFCCVS LCILVAKHTD AGVIQSPRHE VTEMGQEVTL RCKPISGHNS
x58323  MGSWTLCCVS LCILVAKHTD AGVIQSPRHE VTEMGQEVTL RCKPISGHDY
x58329  MGTRLLCWAA LCLLGADHTG AGVSQTPSNK VTEKGKYVEL RCDPISGHTA
x58326  ....LLFWVA FCLLGADHTG AGVSQSPSNK VTEKGKDVEL RCDPISGHTA
x57536  MGTRLLCWAA LCLLGAELTE AGVAQSPRYK IIEKRQSVAF WCNPISGHAT
x58330  ...RLLCCVA ISFWGARLTD TKVTQRPRLL VKASEQKAKM DCVPIKAHSY
x58328  MGSRLLCWVL LCLLGAGPVK AGVTQTPRYL IKTRGQQVTL SCSPISGHRS
x58331  .......... .......... .......... .......... ..........


         >                 < CDR2  >                        100
x58325  MYWYKQDSKK FLKIMFSYNN KELIINETVP N.RFSPKSPD KAHLNLHINS
x57535  MYWYKQDSKK FLKIMFSYNN KELIINETVP N.RFSPKSPD KAHLNLHINS
x58324  LFWYRQTMMR GLELLIYFNN NVPIDDSGMP EDRFSAKMPN ASFSTLKIQP
x58323  LFWYRQTMMR GLELLIYFNN NVPIDDSGMP EDRFSAKMPN ASFSTLKIQP
x58329  LYWYRQSLGQ GPEFLIYFQG TGAADDSGLP NDRFFAVRPE GSVSTLKIQR
x58326  LYWYRQSLGQ GLEFLIYFQG NSAPDKSGLP SDRFSAERTG GSVSTLTIQR
x57536  LYWYQQILGQ GPKLLIQFQN NGVVDDSQLP KDRFSAERLK GVDSTLKIQP
x58330  VYWYRKKLEE ELKFLVYFQN EELIQKAEII NERFLAQCSK NSSCTLEIQS
x58328  VSWYQQTPGQ GLQFLFEYFS ETQRNKGNFP .GRFSGRQFS NSRSEMNVST
x58331  .......... .......... .....KGEVP DG.YNVSRLK KQNFLLGLES


        101         <   CDR3     >         135
x58325  LELGDSAVYF CASSQGRSSN QPQHFGDGTR LSILE
x57535  LELGDSAVYF CASSQGRSSN QPQHFGDGTR LSILE
x58324  SEPRDSAVYF CASSLTGFGD SPLHFGNGTR LTVTE
x58323  SEPRDLAVYF CASSL.ELGT EAF.FGQGTR LTVVE
x58329  TERGDSAVYL CASSWSGRDG DTQYFGPGTR LTVLE
x58326  TQQEDSAVYL CASSREF..S SEQFFGPGTR LTVLE
x57536  AKLEDSAVYL CASS..LLRS WEQFFGPGTR LTVLE
x58330  TESGDTALYF CASSKPGTSP NEQFFGPGTR LTVLE
x58328  LELGDSALYL CASRSSG.GI YEQYFGPGTR LTVTE
x58331  AAPSQTSVYF CASREFR... DTQYFGPGTR LTVLE
```

Figure 2.5 PILEUP of myelin basic protein specific human TCR Vβ sequences (without sequenced Vα partners) [Group 2]. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: @murine.alpha.pigeoncytochromec

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 murine.alpha.pigeoncytochromec  MSF: 132  Type: P  August 25, 1995
17:27  Check: 2963 ..

 Name: m20876          Len:   132  Check: 9914  Weight:  1.00
 Name: m20875          Len:   132  Check: 3049  Weight:  1.00

//

        1                                                  <  CDR1
m20876  .......LTQ VCCASQLGLK EQQVQQSPAS LVLQEAENAE LQCSF SIFTNQ
m20875  MKSLLSSLLG LLCTQVCWVK GQQVQQSPAS LVLQEGENAE LQCNF STSLNS

          >                < CDR2 >                         100
m20876   VQWFYQRPG GRLVSLLYNP SGTKQSGRLT STTVIKERRS SLHISSSQIT
m20875   MQWFYQRPG GSLVSLFYNP SGTKHSGRLT STTVIKERRS SLHISSSQTT

        101    <    CDR3   >           132
m20876  DSGTYLCA MEPTNTGK LTFGDGTVLTVKPNIQ
m20875  DSGTYLCA LVSSNTNK VVFGTGTRLQVLPNIQ
```

Figure 2.6 PILEUP of murine TCR Vα sequences (with Vβ partners) from pigeon cytochrome c restricted clones [Group 2]. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: @murine.beta.pigeon.cytochromec

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 murine.beta.pigeon.cytochromec.msf  MSF: 116  Type: P  August 16,
1995 17:31  Check: 1737 ..

 Name: m20878           Len:   116  Check: 2204  Weight:  1.00
 Name: m20877           Len:   116  Check: 9533  Weight:  1.00

//

        1                          <  CDR1   >                <
m20878  LMNTKITQSP RYLILGRANK SLECQQHLGH NAMYWYKQSA EKPPELMFLY
m20877  LMNTKITQSP RYLILGRANK SLECEQHLGH NAMYWYKQSA EKPPELMFLY

        CDR2 >                                           <   CDR3
m20878  NLKQLIRNET VPSRFIPECP DSSKLLLHIS AVDPEDSAVY FCASSQEGGA
m20877  NLKQLIRNET VPSRFIPECP DSSKLLLHIS AVDPEDSAVY FCASSQDQNS

             >        116
m20878  NYAEQFFGPG TRLTVL
m20877  DYT...FGSG TRLLVI
```

Figure 2.7 PILEUP of murine TCR Vβ sequences from pigeon cytochrome c restricted clones [Group 2]. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: class2.pair.alpha.pep{*}

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 class2.pair.alpha.pep.pileup.msf  MSF: 153  Type: P  November 26,
1995 21:33  Check: 6933 ..
//
        1                                                    <   CDR1
m97710  ......... ......... DSVTQMEGPV TLSEEAFLTI NCTYTATGYP
m97706  ......... ......... DSVTQMEGPV TLSEEAFLTI NCTYTATGYP
m86361  MLLLLVPAFQ VIFTLGGTRA QSVTQLDSQV PVFEEAPVEL RCNYSSSVSV
x63455  MLLLLVPVLE VIFTLGGTRA QSVTQLGSHV SVSEGALVLL RCNYSSSVPP
 cl-1a  ......... ......... QSVTQLDSQV PVFEEAPVEL RCNYSSSVSV
m97708  ......... ......... IQVEQSPPDL ILQEGANSTL RCNFSDS.VN
m97720  ......... ......... IQVEQSPPDL ILQEGANSTL RCNFSDS.VN
m97722  ......... ......... IQVEQSPPDL ILQEGANSTL RCNFSDS.VN
m97724  ......... ......... NEVEQSPQNL TAQEGEFITI NCSYSVG.IS
m97718  ......... ......... QQVKQSPQSL IVQKGGISII NCAYENTAFD
m97704  ......... ......... QQVKQSPQSL IVQKGGISII NCAYENTAFD
m97714  ......... ......... ENVEQHPSTL SVQEGDSAVI KCTYSDSASN
m97712  ......... ......... DQVTQSPEAL RLQEGESSSL NCSYTVSGLR
m97716  ......... ......... ......... ......... .........

          >                 < CDR2  >                        100
m97710  SLFWYVQYPG EGLQLLLKAT KADD..KGSN KGFEATYRKE TTSFHLEKGS
m97706  SLFWYVQYPG EGLQLLLKAT KADD..KGSN KGFEATYRKE TTSFHLEKGS
m86361  YLFWYVQYPN QGLQLLLKYL SGSTLVKGIN .GFEAEFNKS QTSFHLRKPS
x63455  YLFWYVQYPN QGLQLLLKYT SAATLVKGIN G.FEAEFKKS ETSFHLTKPS
 cl-1a  YLFWYVQYPN QGLQLLLKYL SGSTLVKGIN G.FEAEFNKS QTSFHLRKPS
m97708  NLQWFHQNPW ..GQLINLFY IPSGTKQNGR ..LSATTVAT ERYSLLYISS
m97720  NLQWFHQNPW ..GQLINLFY IPSGTKQNGR ..LSATTVAT ERYSLLYISS
m97722  NLQWFHQNPW ..GQLINLFY IPSGTKQNGR ..LSATTVAT ERYSLLYISS
m97724  ALHWLQQHPG ..GGIVSLFM LSSGKKKHGR ..LIATINIQ EKHSSLHITA
m97718  YFPWYQQFPG KGPALLIAIR PDVSEKKEGR ..FTISFNKS AKQFSLHIMD
m97704  YFPWYQQFPG KGPALLIAIR PDVSEKKEGR ..FTISFNKS AKQFSLHIMD
m97714  YFPWYKQELG KRPQLIIDIR SNVGEKKDQR ..IAVTLNKT AKHFSLHITE
m97712  GLFWYRQDPG KGPEFLFTLY SAGEEKEKER ..LKATLTK. .KESFLHITA
m97716  ...WYRQDCR KEPKLLMSVY SSGNE..DGR ..FTAQLNRA SQYISLLIRD

        101         <     CDR3      >                      150
m97710  VQVSDSAVYF CALSD..Q.N GN..KLVFGA GTILRVKSYI QNPDPAVYQL
m97706  VQVSDSAVYF CALSR..S.S GSARQLTFGS GTQLTVLPDI QNPD......
m86361  VHISDTAEYF CAVRF..SLQ GGSEKLVFGK GTKLTVNPY. ..........
x63455  AHMSDAAEYF CAV...SESP FGNEKLTFGT GTRLTIIPNI QNPDPAVYQL
 cl-1a  VHISDTAEYF CAV.....SE TGGFKTIFGA GTRLFVKANI QNPDP.....
m97708  SQTTDSGVYF CAVS...ASG AGSYQLTFGK GTKLSVIPNI QNPDPAVYQL
m97720  SQTTDSGVYF CAVG...FSD GQ..KLLFAR GTMLKVDLNI QNPDPAVYQL
m97722  SQTTDSGVYF CAVG...DYG QS...FVFGP GTRLSVLPYI QKPDPAV...
m97724  SHPRDSAVYI CAV.....YS QGAQKLVFGQ GTRLTINPNI Q.........
m97718  SQPGDSATYF CAASKRGAGG TSYGKLTFGQ GTILTVHPNI QNP.......
m97704  SQPGDSATYF CAASR.....  ..ANNLFFGT GTRLTVIPYI QNPDPAVYQL
m97714  TQPEDSAVYF CAASR...GG GATNKLIFGT GTLLAVQPNI QNPDPAV...
m97712  PKPEDSATYL CAVP...TTT DSWGKLQFGA GTQVVVTPDI QNP.......
m97716  SKLSDSATYL CVVILPNNAG ...NMLTFGG GTRLMVKPHI QNPDPAV...
```

Figure 2.8 PILEUP of human TCR Vα sequences (with sequenced Vβ partners) from Class II restricted clones. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: @class2.beta.paired.pep

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 class2.beta.paired.pep.msf  MSF: 321  Type: P  August 14, 1995
11:48
//
       1                                              <   CDR1  >
m97723  ..................SAVISQKPSRDICQRGTSLTIQCQ.V .DSQVTM
m97713  ..................SAVISQKPSRDICQRGTSLTIQCQ.V .DSQVTM
m97719  ..................GAVVSQHPSWVICKSGTSVKIECRS. LDFQAT.
m97717  ..................GAVVSQHPSWVICKSGTSVKIECRS. LDFQAT.
m97721  ..................GAVVSQHPSRVICKSGTSVKIECRS. LDFQAT.
m97707  ..................ETGVTQSPTHLIKTRGQQVTLRC.SS QSGHNT.
m86362  PGLLCWVLLCLLGAGPVDAGVTQSPTHLIKTRGQHVTLRC.SP ISGHKS.
m97709  ..................EAGVTQSPTHLIKTRGQQATLRC.SP ISGHTS.
m97705  ..................KAGVTQTPRYLIKTRGQQVTLSC.SP ISGHRS.
x63456  IRLLCRVAFCFLAVGLVDVKVTQSSRYLVKRTGEKVFLEC.VQ DMDHEN.
m97715  ..................KAGVTQTPRYLIKTRGQQVTLSC.SP ISGHRS.
m97725  ..................DGGITQSPKYLFRKEGQNVTLSC.EQ NLNHDA.
z26594  IGLLCCAALSLLWAGPVNAGVTQTPKFQVLKTGQSMTLQC.AQ DMNHEY.
m97711  ..................NAGVMQNPRHLVRRRGQEARLRC.SP MKGHSH.

                           < CDR2 >
m97723  .MFWYRQQPGQSLTLI. ATANQGSEATYESGF .VIDKFPISRPNLTFSTL
m97713  .MFWYRQQPGQSLTLI. ATANQGSEATYESGF .VIDKFPISRPNLTFSTL
m97719  TMFWYRQFPKQSLMLM. ATSNEGSKATYEQGV .EKDKFLINHASLTLSTL
m97717  TMFWYRQFPKQSLMLM. ATSNEGSKATYEQGV .EKDKFLINHASLTLSTL
m97721  TMFWYRQFPKQSLMLM. ATSNEGSKATYEQGV .EKDKFLINHASLTLSTL
m97707  .VSWYQQALGQGPQFI. .FQYYREEENGRGN. FP.PRFSGLQFPNYSSEL
m86362  .VSWYQQVLGQGPQFI. .FQYYEKEERGRGN. FP.DRFSARQFPNYSSEL
m97709  .VYWYQQALGLGLQFL. .LWYDEGEERNRGN. FP.PRFSGRQFPNYSSEL
m97705  .VSWYQQTPGQGLQFL. .FEYFSETQRNKGN. FP.GRFSGRQFSNSRSEM
x63456  .MFWYRQDPGLGLRLI. YFSYDVKMKE.KGD. IPEG.YSVSREKKERFSL
m97715  .VSWYQQTPGQGLQFL. .FEYFSETQRNKGN. FP.GRFSGRQFSNSRSEM
m97725  .MYWYRQDPGQGLRLI. YYSHIVNDFQ.KGD. IAEG.YSVSREKKESFPL
z26594  .MSWYRQDPGMGLRLIH .YSVGAGITD.QGE. VPNG.YNVSRSTTEDFPL
m97711  .VYWYRQLPEEGLKFM. VYLQKENIID.ESGM .PKERFSAEFPKEGPSIL

                        <       CDR3         >
m97723  TVSNMSPEDSSIYLC..S..  .VEDRDRVYNE. QFFGPGTRLTVLEDLKNV
m97713  TVSNMSPEDSSIYLC..S..  .VAGQGT..DE. QYFGPGTRLTVTEDLKNV
m97719  TVTSAHPEDSSFYIC..S.A ...SKTGR.GE. QYFGPGTRLTVTEDLKNV
m97717  TVTSAHPEDSSFYIC..S..  ..ASRASSYNE. QFFGPGTRLTVLEDLKNV
m97721  TVTSAHPEDSSFYIC..S.A ...LGTGS..GY .TFGSGTRLTVVEDLNKV
m97707  NVNALELDDSALYLCAS... SLMGDRG..SPL .HFGNGTRLTVTEDLNKV
m86362  NVNALLLGDSALYLCAS... SFPGNRGYWYGY .TFGSGTRLTVVED....
m97709  NVNALELEDSALYLCAS... SL.GQGT..YE. QYFGPGTRLTVTEDLKNV
m97705  NVSTLELGDSALYLCAS... SKEPRGNQ..P. QHFGDGTRLSILEDLNKV
x63456  ILESASTNQTSMYLCAS... SSTGLPY...GY .TFGSGTRLTVVEDLNKV
m97715  NVSTLELGDSALYLCAS... SSTTSGSN..E. QYFGPGTRLTVTEDLKNV
m97725  TVTSAQKNPTAFYLCAS... SLDSWDT..... QYFGPGTRLTVLEDLKNV
z26594  RLLSAAPSQTSVYFCAS... R.DFLSG...E. QYFGPGTRLTVTEDLKNV
m97711  RIQQVVRGDSAAYFCAS... SPTVS....YE. QYFGPGTRLTVTEDLKNV
```

Figure 2.9 PILEUP of human TCR Vβ sequences (with Vα partners) from Class II restricted clones. The first column refers to the EMBL sequence accession numbers in the alignments.

```
PileUp of: @class1.pair.alpha.pep

 Symbol comparison table: GenRunData:pileuppep.cmp  CompCheck: 1254

                 GapWeight: 3.000
            GapLengthWeight: 0.100

 class1.pair.alpha.pileup.msf  MSF: 283  Type: P  November 26, 1995
21:48  Check: 6003 ..

 //

        1                                                        <
x74390  ......... ......... ......... ......... .........
s60795  ....RTMKTF AGFSFLFLWL QLDCMSRGED VEQSLFLSVR EGDSSVINCT
x74394  ......... ......... ......... ......... .........
x74396  ......... ......... ......... ......... .........
l38312  ......... ......... .......QEV TQIPAALSVP EGENLVLNCS
s69564  ......... ......... ......... ......... .........
m81774  MMKSLRVLLV ILWLQLSWVW SQQK....EV EQNSGPLSVP EGAIASLNCT
s69559  ......... ......... ......... ......... .........
m15565  MAMLLGASVL ILWLQPDWVN SQQKNDDQQV KQNSPSLSVQ EGRISILNCD
s60793  ......... ......... ......... ......... .........
s69563  ......... ......... ......... ......... .........
x74392  ......... ......... ......... ......... .........
s69553  ......... ......... ......... ......... .........

         CDR1    >                 < CDR2 >                100
x74390  ......... ......... ......... ......... .........
s60795  YTDSSSTYLY WYKQEPGAGL QLLTYIFSNM DMKQDQRLTV LLNKKDKHLS
x74394  ......... ......... ......... ......... .........
x74396  ......... ......... ......... ......... .........
l38312  FTDSAIYNLQ WFRQDPGKGL TSLLLIQSSQ REQTSGRLNA SLDKSSGRST
s69564  ......... ......... ......... ......... .........
m81774  YSDRGSQSFF WYRQYSGKSP ELIMSIYSNG DK.EDGRFTA QLNKASQYVS
s69559  ......... ......... ......... ......... .........
m15565  YTNSMFDYFL WYKKYPAEGP TFLISISSIK DKNEDGRFTV FLNKSAKHLS
s60793  ......... ......... ......... ......... .........
s69563  ......... ......... ......... ......... .........
x74392  ......... ......... ......... ......... .........
s69553  ......... ......... ......... ......... .........

        101              <   CDR3      >                    150
x74390  ......... .....CAVKG IS.GGSYIPT FGRG.TSLIV HPYIQ.....
s60795  LRIADTQTGD SAIYFCAEKN SKLGGSYIPT FGRG.TSLIV HPYIQNP...
x74394  ......... ......CAVN TG.GFKTI.. FGAG.TRLFV KANIQ.....
x74396  ......... ......CAVN KG.N.DMR.. FGAG.TRLTV KPNIQ.....
l38312  LYIAASQPGD SATYLCAALD TG.RRALT.. FGSG.TRLQV QPNIQ.....
s69564  ......... ..EYFCAV.. GATGNQFY.. FGTG.TSLTV IPNIQNPD..
m81774  LLIRDSQPSD SATYLCAVYH SGSARQLT.. FGSG.TQLTV LPDIQNPDPA
s69559  ......... ..TYFCAASK GGSQGNLI.. FGKG.TKLSV KPNIQNPD..
m15565  LHIVPSQPGD SAVYFCAAKG AGTASKLT.. FGTG.TRLQV TLDIQNPDPA
s60793  ......... ......AASI TNSGYALN.. FGKG.TSLLV TPHIQNP...
s69563  ......... ..MYYCALIP GG..QKLL.. FARG.TMLKV DLNIQNPD..
x74392  ......... .......CAA RLLTYKYI.. FGTG.TRLKV LANIQ.....
s69553  ......... ..MYFCAYRG LGVVLQTSSS LELALCLLSS QVHIQNPD..
```

Figure 2.10 PILEUP of human TCR Vα sequences (with Vβ partners) from Class I restricted clones [Group 4]. The first column refers to the EMBL sequence accession numbers in the alignments.

```
Pileup of :  @class1prb.msf
class1.pair.beta.pileup.msf MSF: 314  Type: P  July 25, 1995 19:32
Check: 6520 ..


//

        1                                                    <  CDR1
s69566  .......... .......... .......... .......... ..........
m15564  .......... .......... .GVSQNPRHN ITKRGQNVTF RCDPISEHNR
s69556  .......... .......... .......... .......... ..........
x74397  .......... .......... .......... .......... ..........
m81773  MGSWTLCCVS LCILVAKHTD AGVIQSPRHE VTEMGQEVTL RCKPISGHDY
x74395  .......... .......... .......... .......... ..........
x74391  .......... .......... .......... .......... ..........
s69571  .......... .......... .......... .......... ..........
s69569  .......... .......... .......... .......... ..........
s60794  .MSISLLCCA AFPLLWAGPV NAGVTQTPKF RILKIGQSMT LQCTQDMNHN
x74393  .......... .......... .......... .......... ..........
s60796  .......... .......... .......... .......... ..........
m24089  MGTRLLFWVA FCLLGADHTG AGVSQSPSNK VTEKGKDVEL RCDPISGHTA


         >               < CDR2 >                            100
s69566  .......... .......... .......... .......... ..........
m15564  LYWYRQTLGQ GPEFLTYFQN EAQLEKSRLL SDRFSAERPK GSFSTLEIQR
s69556  .......... .......... .......... .......... ..........
x74397  .......... .......... .......... .......... ..........
m81773  LFWYRQTMMR GLELLIYFNN NVPIDDSGMP EDRFSAKMPN ASFSTLKIQP
x74395  .......... .......... .......... .......... ..........
x74391  .......... .......... .......... .......... ..........
s69571  .......... .......... .......... .......... ..........
s69569  .......... .......... .......... .......... ..........
s60794  YMYWYRQDPG MGLKLIYYSV GAGITDKGEV PNGYNVSRST TEDFPLRLEL
x74393  .......... .......... .......... .......... ..........
s60796  .......... .......... .......... .......... ..........
m24089  LYWYRQSLGQ GLEFLIYFQG NSAPDKSGLP SDRFSAERTG GSVSTLTIQR


        101          <      CDR3     >                       150
s69566  .......MYL CASSLVVWDR GGNQPQHFGD GTRLSILEDL NKV.......
m15564  TEQGDSAMYL CASSLA.... GLNQPQHFGD GTRLSILEDL NKVFPPEVAV
s69556  ........YF CASRPTITVP YSNQPQHFGD GTRLSILEDL NKV.......
x74397  .......... CASSYS.AA. .SGHGYTFGS GTRLTVVEDL ..........
m81773  SEPRDSAVYF CASSLT.TER IPVYGYTFGS GTRLTVVEDL NKVFPPEVAV
x74395  .......... CASSTG.QGW GS.....FGS GTRLTVVEDL ..........
x74391  .......... CASR...S.. .SREEQFFGP GTRLTVLEDL ..........
s69571  .......LYL CASS...QDL LSWDEQFFGP GTRLTVLEDL KNV.......
s69569  ......MYFC AAGE...TSG VSYNEQFFGP GTRLTVLEDL KNV.......
s60794  AAPSQTSVYF CASSYA.GAL LNTDTQYFGP GTRLTVLEDL K.........
x74393  .......... ........CS AFDGEAFFGQ GTRLTVVEDL ..........
s60796  .......... ..TSEP.GHP ILSYEQYFGP GTRLTVTEDL K.........
m24089  TQQEDSAVYL CASSLV.SGR AG.DTQYFGP GTRLTVL... ..........
```

Figure 2.11 PILEUP of Class I restricted TCR Vβ sequences (with sequenced Vα partners) [Group 4]. The first column refers to the EMBL sequence accession numbers in the alignments.

Another way of characterising a group of sequences is by calculating the percentage of residues that are 'absolutely positively conserved' (APC) throughout the alignment. The MPI and APC values of each of these sets in the four groups were calculated (Table 2.2 and 2.3) and are shown in graphical form in Figures 2.12 and 2.13.

The lower the MPI and APC values are for the group, the greater the sequence divergence within the group. It was found that both APC and MPI values decreased as the peptide and MHC molecule was made random in both TCR Vα and Vβ sequences. Group 1 which had different TCR sets with the same peptide and MHC combination showed the highest values of both APC and MPI. The order of similarity increased from Group 1 > Group 2 > Group 3 > Group 4. Group 4 which has TCR sets with random peptide and MHC had low values of APC and MPI suggesting that sequences in the different sets in this group are far more diverse than those in other groups.

After calculating the overall diversity in these sequences in the four groups the overall similarity in these sequences was calculated by a GCG program PLOTSIMILARITY.

2.4.2 PLOTSIMILARITY

PLOTSIMILARITY plots the running average of the similarity among the sequences in a multiple sequence format using a user specified sliding window of comparison. The window of comparison is moved along all sequences, one position at a time and the average similarity over the entire window is plotted at the middle

Group	Peptide	MHC	APC (%)	MPI (%)
Group 1	***same peptide***	***same MHC***		
	Flu matrix peptide	HLA-A2.1	16.66	22.22
	Human dust mite	DR1	18.66	25.33
	Lcv	H-2^b	35.71	46.42
	Tumour necrosis protein	H-$2K^b$	50	64.28
Group 2	***same peptide***	***random MHC***		
	Myelin basic protein	DR family specific	9.16	27.5
	Betv	DR family specific	20.83	31.66
	Pigeon cytochrome c	I-E^k, I-E^b	68.93	68.93
Group 3	***random peptide***	***allo MHC***		
	Varied (Vα sequences with Vβ partners)	HLA-DR1	9.16	19.16
	Varied	HLA-B27	9.09	28.71
Group 4	***random peptide***	***random MHC***		
	Varied (Vα sequences with no Vβ partners)	Class I	0	4.76
	Varied (Vα sequences with no Vβ partners)	Class II	0	5.55
	Varied (Flu matrix peptide + HA)	HLA-DR1, HLA-A2	5	9.165
	Varied (Vα sequences with Vβ partners)	Class I	5.55	16.66

Table 2.2 APC and MPI values of TCR Vα sequences.

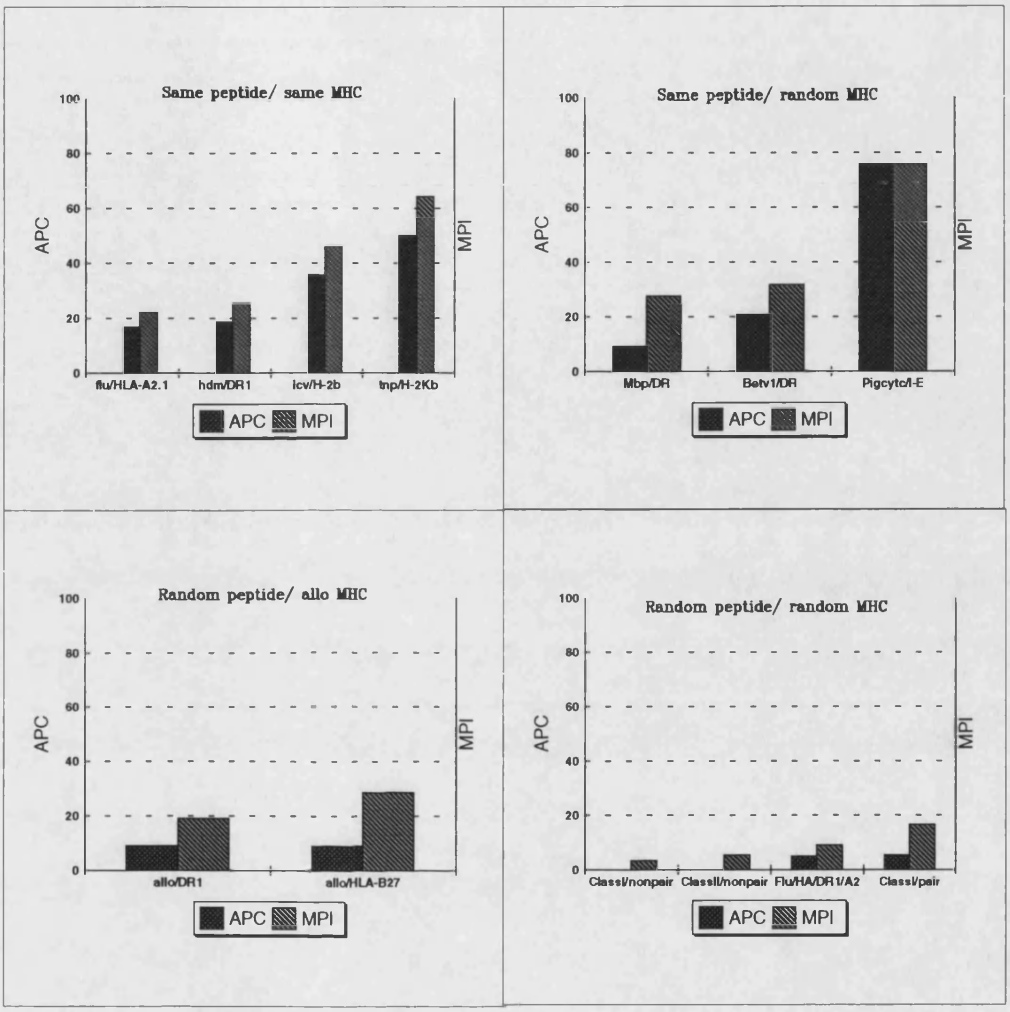


Figure 2.12 APC and MPI values of TCR V alpha sequences shown as histograms

Group	Peptide	MHC	APC (%)	MPI (%)
Group 1	***same peptide***	***same MHC***		
	Human dust mite	HLA-DR1	19.23	19.23
	Tumour necrosis protein	H-2K^b	0	37.5
	Flu matrix peptide	HLA-A2.1	29.41	41.17
	Lcv	H-2^b	50	64.28
Group 2	***same peptide***	***random MHC***		
	Betv	DR family specific	4.32	22.58
	Myelin basic protein	DR family specific	9.165	17.5
	Pigeon cytochrome c	I-E^k, I-E^b	88.79	88.79
Group 3	***random peptide***	***allo MHC***		
	Varied (Vβ sequences with Vα partners)	HLA-DR1	13.33	30
	Varied (Vβ sequences with no Vα partners)	HLA-DR1	15.38	33.84
Group 4	***random peptide***	***random MHC***		
	Varied (Flu matrix peptide + HA)	HLA-DR1, HLA-A2	0	5.88
	Varied (Vβ sequences with Vα partners)	Class I	5	15
	Varied (Vβ sequences with no Vα partners)	Class II	4.16	25
	Varied (Vβ sequences with no Vα partners)	Class I	13.89	28.57

Table 2.3 APC and MPI values of TCR Vβ sequences

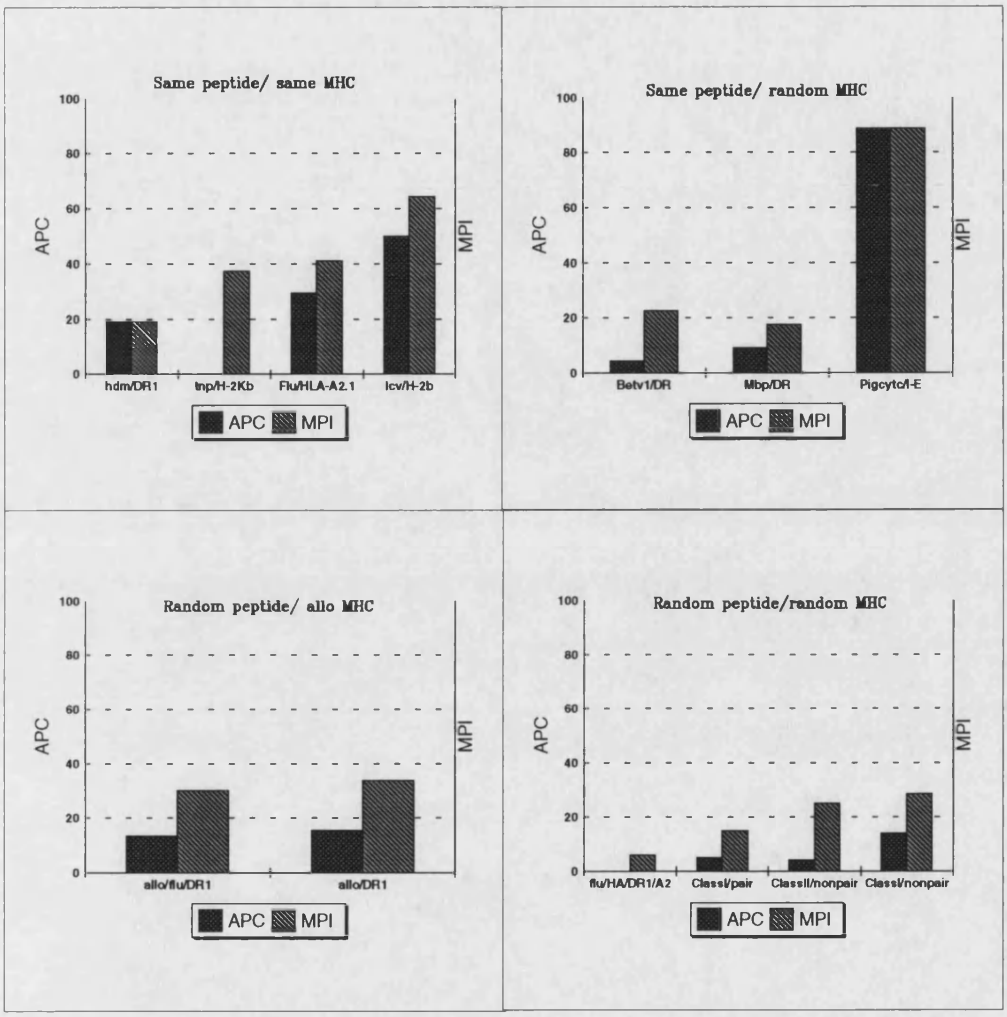


Figure 2.13 APC and MPI values of TCR V beta sequences shown as histograms.

position of the window. The average similarity at a position in an alignment is the arithmetic average of the scores of all possible pairwise symbol comparisons among the sequence symbols at that position. The average similarity across the entire alignment is the sum of the separate window similarities divided by the number of windows and it is plotted as a dotted line.

A measure of the level of identity among all the sequences in a multiple sequence alignment can be plotted by using the -IDEntity command line option while a measure of the level of conservation in a profile created from a multiple sequence alignment can be plotted by using the -PROfile command line option.

The default file plotsimpep.cmp has a value of 1.5 for perfect symbol matches and values less than 1.5 (depending upon the evolutionary distance) for non-matches. FETCH can be used to copy these files and they can then be modified to suit one's needs.

TCR Vα and Vβ multiple sequence sets in the four groups were aligned by PILEUP. The average similarity across the various alignments was calculated (Table 2.4) by PLOTSIMILARITY.

The trend seen in the case of APC and MPI values is not duplicated across the spectrum. The sequence sets in Group 2 were found to have consistently high values and these are comparable to the values seen in Group 1. The similarity scores exhibited by Groups 3 and 4 were much lower than the values seen in Groups 1 and 2 (Figure 2.14). Additionally the similarity across the three CDRs in the TCR Vα (Table 2.5) and Vβ sequence sets (Table 2.6) in Groups 2, 3 and 4 was calculated. Similarity plots of CDRs in TCR Vα (Figure 2.15) and Vβ (Figure 2.16) show a

Group	Peptide	MHC	Vα	Vβ
Group 1	***same peptide***	***same MHC***		
	Flu matrix peptide	HLA-A2.1	0.625	0.7
	Human dust mite	HLA-DR1	0.4	0.5
	Tumour necrosis protein	H-2K^b	0.1	0.25
	Lcv	H-2^b	-	0.925
Group 2	***same peptide***	***random MHC***		
	Betv	DR family specific	0.775	0.4165
	Myelin basic protein	DR family specific	0.525	0.725
	Pigeon cytochrome c	I-E^k, I-E^b	1.1	1.375
Group 3	***random peptide***	***allo MHC***		
	Varied (Vα sequences with Vβ partners)	HLA-DR1	0.575	0.3
	Varied	HLA-B27	0.25	-
	Varied	HLA-DR1	-	0.325
Group 4	***random peptide***	***random MHC***		
	Varied (Vβ sequences with Vα partners)	Class I	0.1	0.1
	Varied (Vα sequences with no Vβ partners)	Class II	-	0.2
	Varied (Vβ sequences with no Vα partners)	Class I	0.225	0.15

Table 2.4 Plotsimilarity scores of TCR Vα and Vβ sequences.

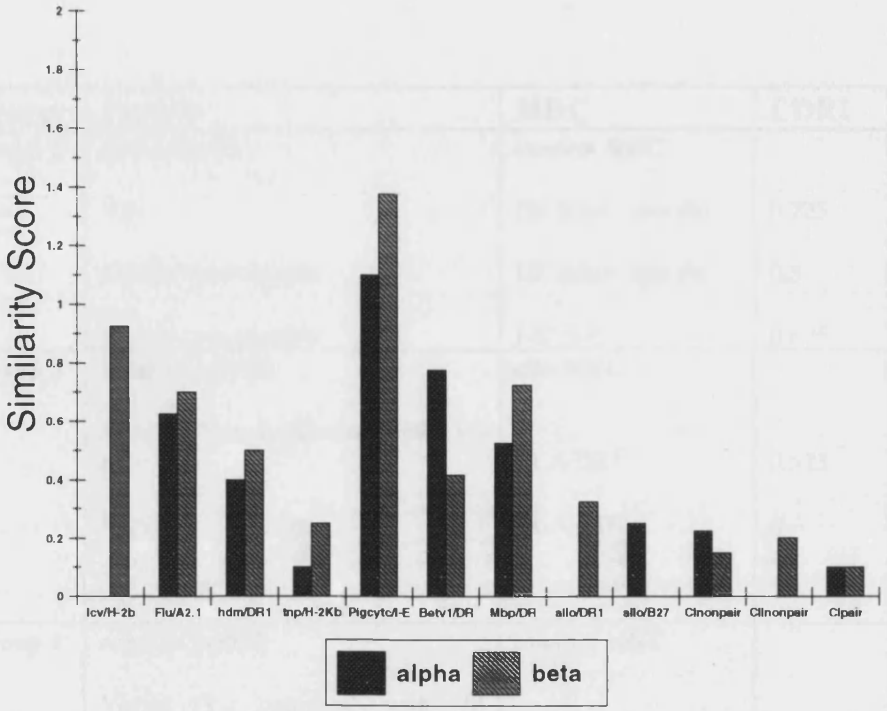


Figure 2.14 Similarity plot of TCR V alpha and V beta sequences

Group	Peptide	MHC	CDR1	CDR2	CDR3
Group 2	***same peptide***	***random MHC***			
	Betv	DR family specific	0.725	0.625	0.325
	Myelin basic protein	DR family specific	0.5	0.35	0.325
	Pigeon cytochrome c	$I\text{-}E^k$, $I\text{-}E^b$	0.675	1.45	0.85
Group 3	***random peptide***	***allo MHC***			
	Varied (Vα sequences with Vβ partners)	HLA-DR1	0.525	0.475	0.575
	Varied	HLA-B27	0	0.175	0.25
Group 4	***random peptide***	***random MHC***			
	Varied (Vα sequences with Vβ partners)	Class I	0.05	0.05	0.325
	Varied (Vα sequences with no Vβ partners)	Class II	0.1	0.1	0.2
	Varied (Vα sequences with no Vβ partners)	Class I	0.025	0.175	0.225

Table 2.5 Plotsimilarity scores of CDRs in TCR Vα sequences.

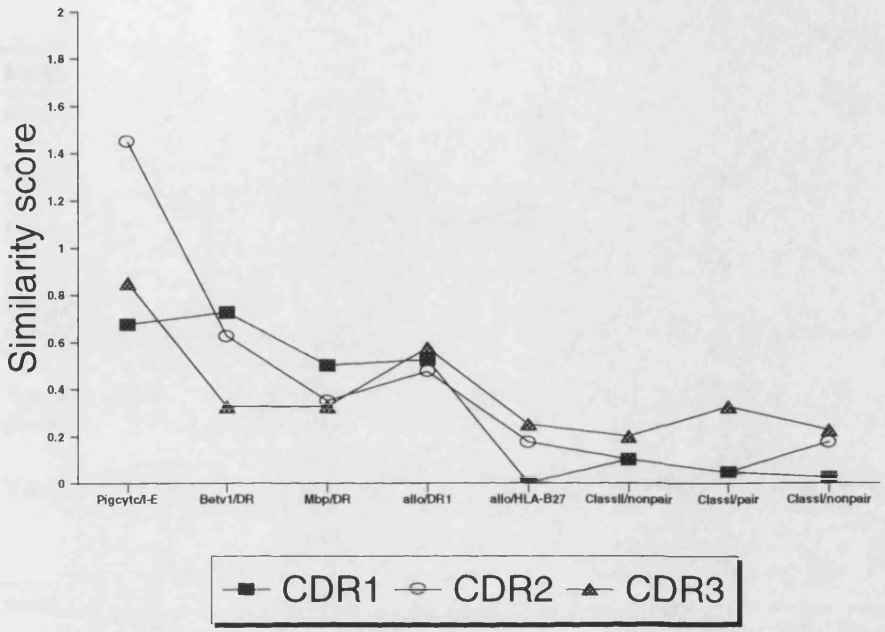


Figure 2.15 Similarity plot of CDRs in TCR V alpha sequences

Group	Peptide	MHC	CDR1	CDR2	CDR3
Group 2	***same peptide***	***random MHC***			
	Betv	DR family specific	0	0.025	0.275
	Myelin basic protein	DR family specific	0.675	0.625	0.625
	Pigeon cytochrome c	$I\text{-}E^k$, $I\text{-}E^b$	1.425	1.5	0.75
Group 3	***random peptide***	***allo MHC***			
	Varied (Vα sequences with Vβ partners)	HLA-DR1	0.475	0.35	0.575
	Varied	HLA-B27	-	-	-
Group 4	***random peptide***	***random MHC***			
	Varied (Vβ sequences with Vα partners)	Class I	0.05	0.05	0.375
	Varied (Vα sequences with no Vβ partners)	Class II	0.3	0.25	0.3
	Varied (Vβ sequences with no Vα partners)	Class I	0.5	-	-

Table 2.6 Plotsimilarity scores of CDRs in TCR Vβ sequences.

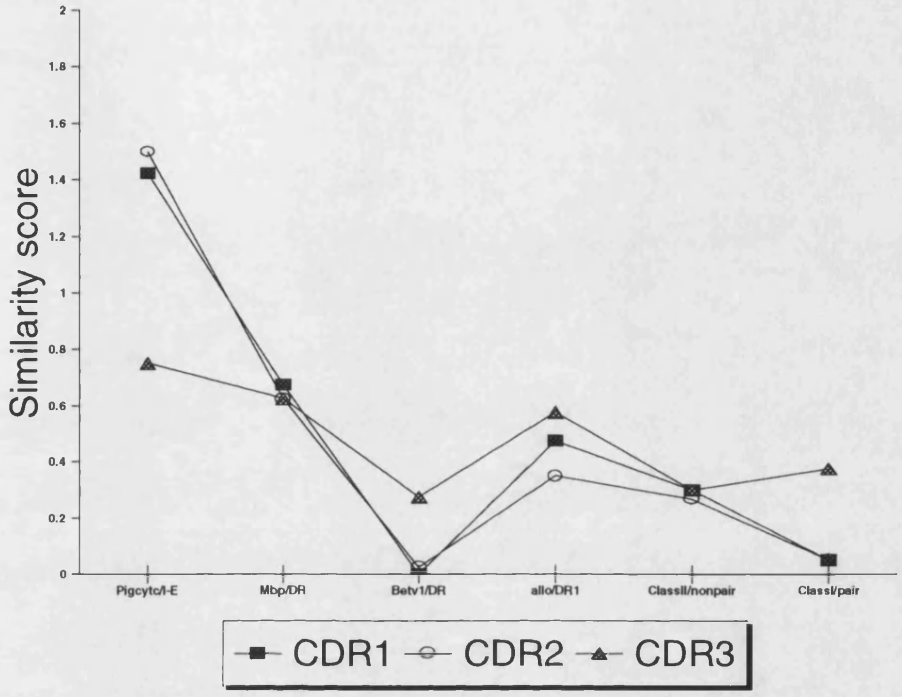


Figure 2.16 Similarity plot of CDRs in TCR V beta sequences

similar pattern to that seen in the case of APC and MPI values.

Generally the CDRs in Group 2 had higher values than the CDRs in Groups 3 and 4. Lower values were seen as the peptide and MHC were made random.

2.4.3 Similarity based on chemical side chain character

While exact conservation of residues at equivalent positions (identity) may not be present, the possibility that similar sidechains at particular positions can maintain specificity at the structural level was investigated. Seventeen out of the twenty amino acids were classified into four groups, based on their chemical similarity.

Group A consisting of hydrophobic amino acids: alanine, isoleucine, leucine, methionine, phenylalanine, valine, tryptophan and tyrosine.

Group B consisting of uncharged hydrophilic amino acids capable of hydrogen bonding: asparagine, glutamine, serine and threonine.

Group C consisting of positively charged amino acids: lysine, arginine and histidine.

Group D consisting of negatively charged amino acids: aspartate and glutamate.

The residues cysteine, proline and glycine are left out from this classification as they do not fall into any specific group and in addition may impose a structural change that cannot readily be predicted.

A simple scoring matrix was derived from the Gribskov table (Table 2.7).

A, I, L, M, F, V, W, Y	N, Q, S, T	K, R, H	D, E	
0.475	- 0.178	- 0.15	- 0.406	A, I, L, M, F, W, V, Y
- 0.178	0.5	0.217	0.425	N, Q, S, T
- 0.15	0.217	0.811	0.233	K, R, H
- 0.406	0.425	0.233	1.25	D, E

Table 2.7 Scoring matrix derived from the Gribskov table.

Gribskov and Burgess rescaled the Dayhoff table (Dayhoff *et al.*, 1978) by dividing each value by the sum of its row and column and normalising to a mean of 0 and standard deviation of 1. Perfect matches were set to have a value of 1.5 and no matches on any row were better than the perfect matches.

The seventeen amino acids were initially aligned into the four groups A, B, C and D. The sum of the values of each row and column in a group was divided by the number of individual elements in that particular group. Each matrix element denotes this average score. It can be noted from the matrix that when a hydrophobic residue (A, I, L, M, F, V, W and Y) is substituted by another hydrophobic residue it is assigned a score of 0.475. However when it is substituted by an uncharged hydrophilic residue (N, Q, S and T), a positively charged residue (K, R and H) or a negatively charged residue (D and E) it is assigned negative scores of -0.178, -0.15 and -0.406 respectively. These scores clearly reflect the difference in the physicochemical properties of the amino acids in the various groups.

Similarly residue substitutions within groups B, C and D are assigned scores of 0.5, 0.811 and 1.25 respectively. The values in this matrix seem to be assigned logically as higher scores are assigned for group substitutions. Substitutions across more diverse groups are assigned lower scores. A high score of 1.25 is seen when an aspartate is substituted by glutamate or *vice versa.*

This matrix was then used to assign scores to each CDR position in the various TCR Vα and Vβ sequence sets in Groups 1 to 4.

For example, at position 1 in CDR1 of mbp specific TCR Vα sequences (Figure 2.4), there are three threonines, one alanine, one glutamate, one

aspartate, one asparagine and three serines. The 3 threonines, 3 serines and 1 asparagine belong to Group B giving a score of 7 × 0.5, the glutamate and aspartate belong to Group D giving a score of 2 × 0.425 while alanine gives a score of -0.178.

The total score at a position is calculated by the formula,

$$\frac{\text{sum of individual scores / number of sequences in that set.}}{\text{maximum score}}$$

In the given example, the total score at position 1 in CDR1 of mbp specific TCR Vα sequences

$$= \frac{3.5+0.85+(-0.178)/\ 10}{0.5} = \frac{0.4172}{0.5} = 0.834$$

In the case of CDR1 and CDR2 position of TCR Vα sequences, higher scores are seen in mbp specific TCR Vα sequences (Group 2) as compared to those in TCR Vα sequences in Class I restricted clones (Group 4) [Figure 2.17]. It should be noted that the Class II paired TCR Vα and Vβ sequences include sequences which are derived from DR1 alloreactive clones and DR1 restricted flu specific clones cl-1 and HA1.7. The CDR1 and CDR2 scores seen in Class II sequences (where there is only a slight variation in the MHC restriction within the sequence set) are comparable to those seen in DR family specific mbp specific TCR Vα and Vβ sequences. As in the case of CDR1 and CDR2, CDR3 scores of the mbp specific TCR Vα sequences (Group 2) show values which are comparable to that seen in Class II restricted TCR Vα sequences.

Even though high scores are seen for residues at CDR3 positions in Group

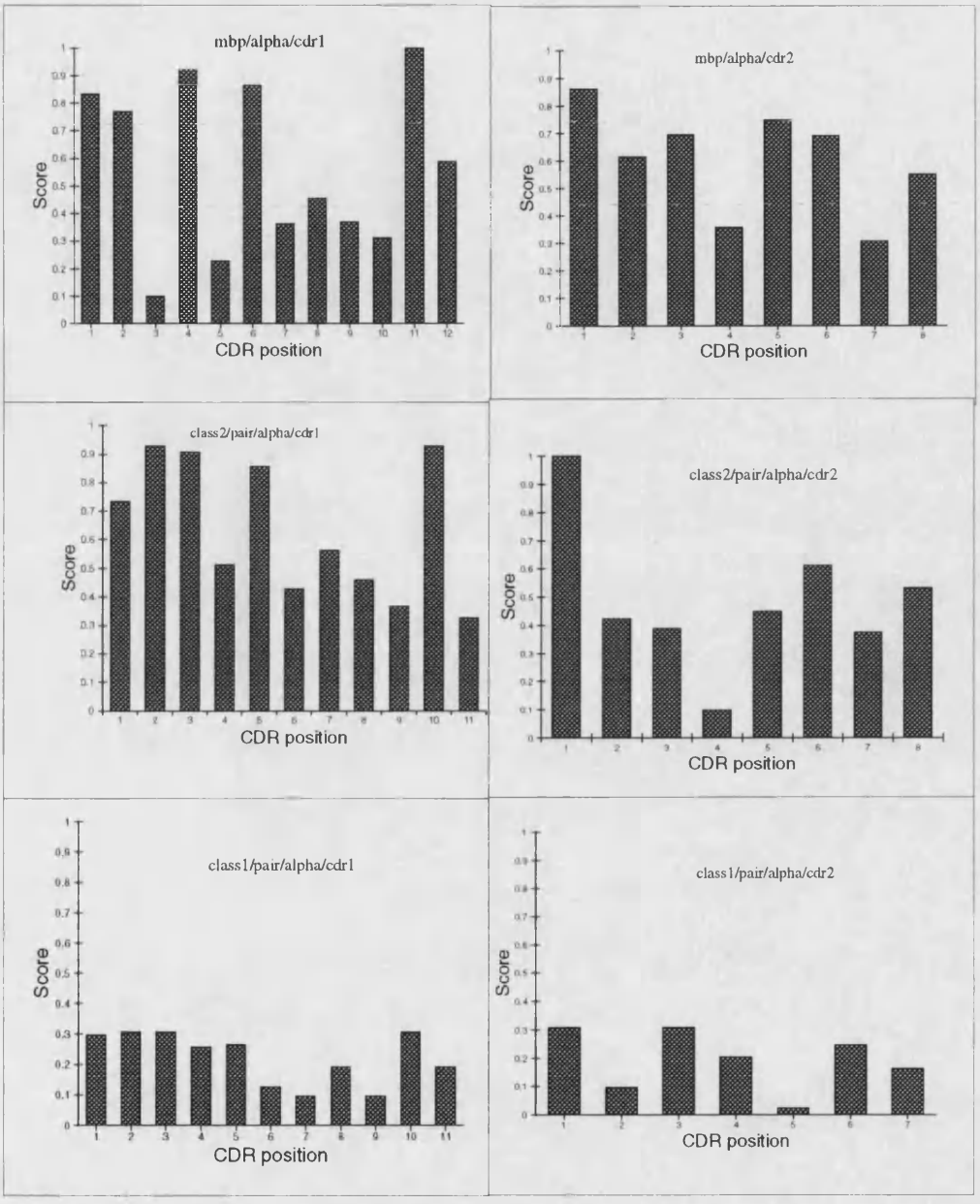


Figure 2.17 CDR1 and CDR2 scores of TCR V alpha sequences from groups 2, 3 and 4.

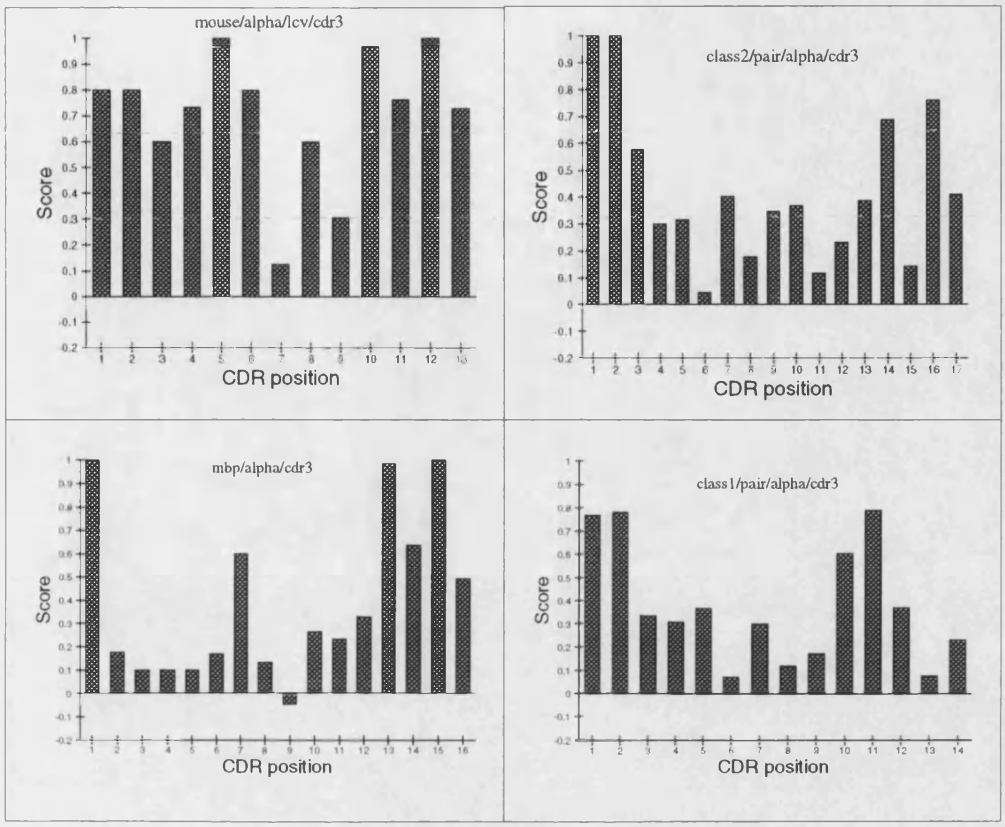


Figure 2.18 CDR3 scores of TCR V alpha sequences from the four groups

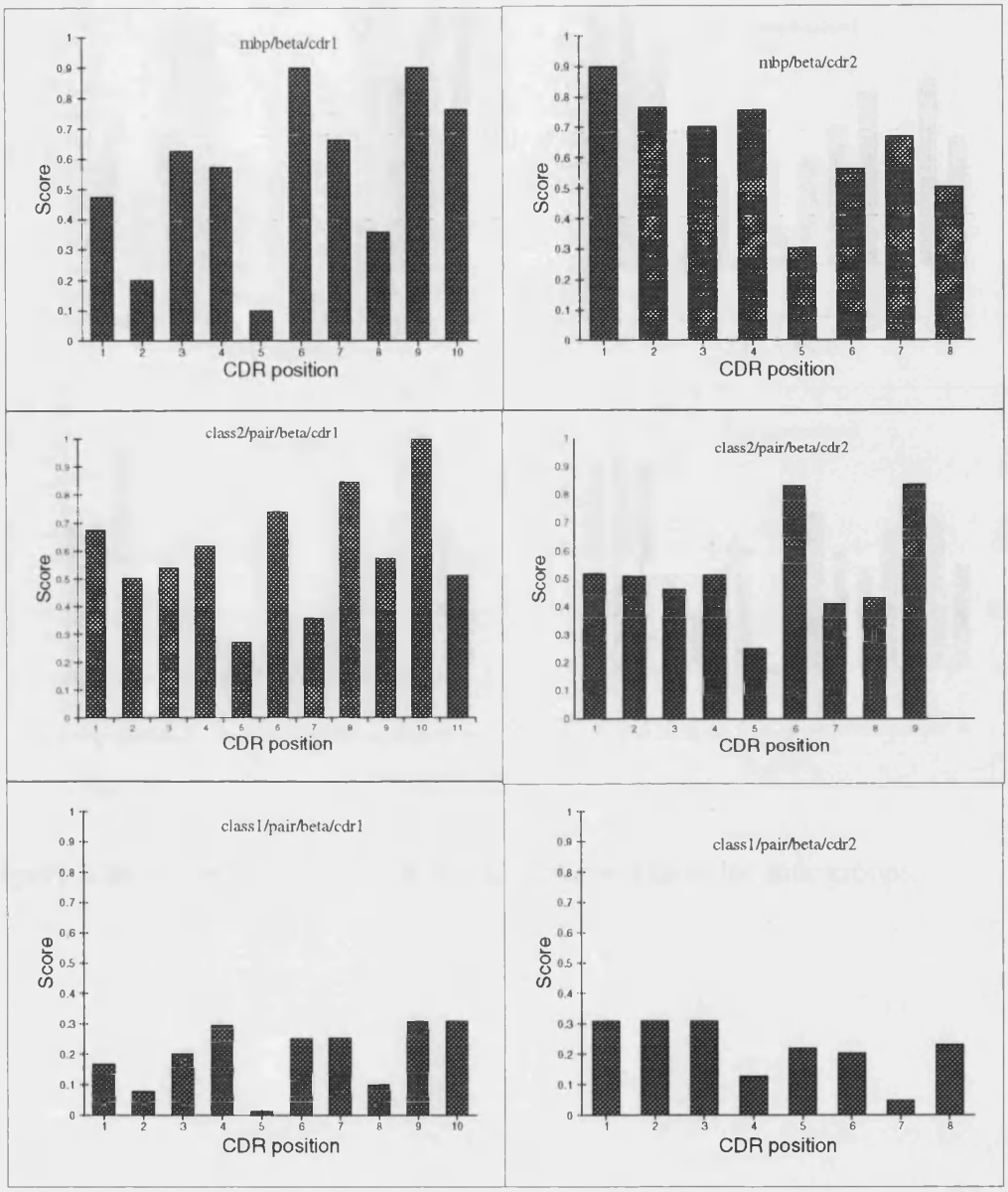


Figure 2.19 CDR1 and CDR2 scores of TCR V beta sequences from groups 2, 3 and 4

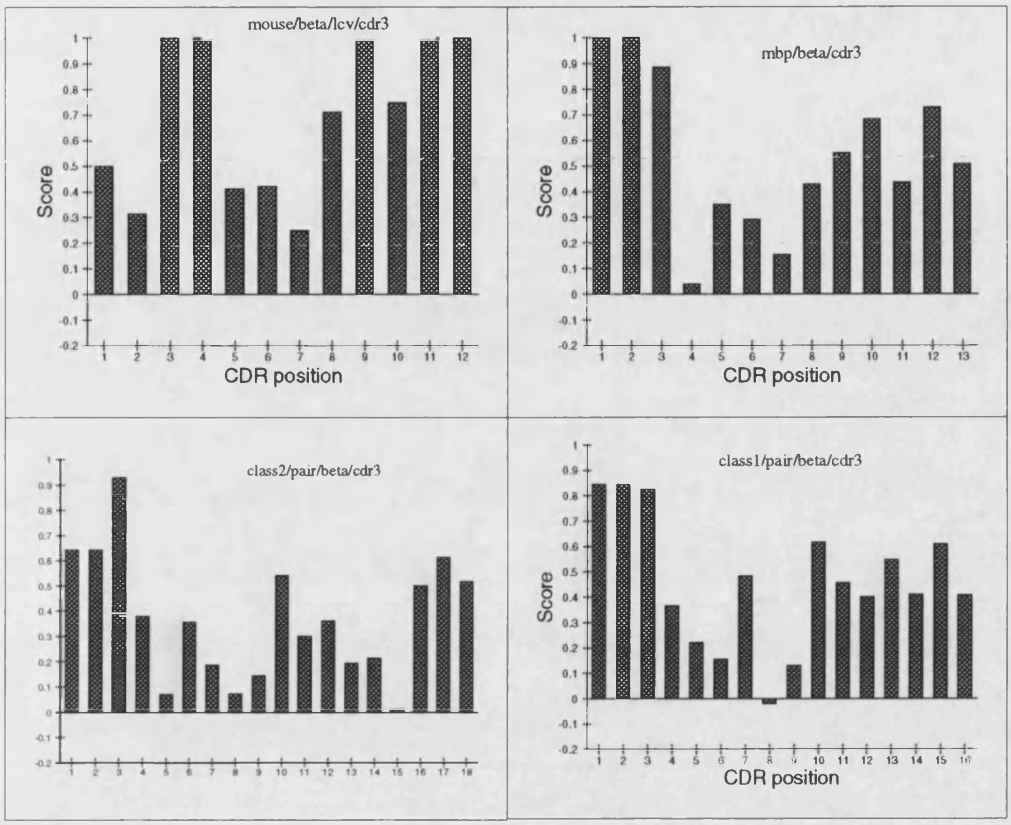


Figure 2.20 CDR3 scores of TCR V beta sequences from the four groups.

1 (lcv specific murine TCR Vα sequences) as compared to those in Group 4 (Class I paired TCR Vα sequences) [Figure 2.18], the similarity scores in the CDR3 positions do not show a great difference across the groups.

High scores are also seen for residues at CDR1 and CDR2 positions in mbp specific TCR Vβ sequences (Group 2) as compared to TCR Vβ sequences in Class I restricted clones (Group 4) [Figure 2.19]. However CDR3 positions in mbp specific DR restricted TCR Vβ sequences and Class I restricted TCR Vβ sequences vary in their scores only at the central CDR positions corresponding to the most accessible portion of the loop (Figure 2.20).

2.5 *Summary of Results*

A large number of TCR Vα and Vβ sequences are increasingly becoming available due to rapid advances in sequencing. These sequences were obtained via anonymous ftp from the various protein and DNA databanks. The sequences were then aligned and assorted depending on their MHC restriction and peptide specificity.

Analysis of the TCR Vα and Vβ gene sequences showed a high variability in the TCR sequences as compared to antibodies. In addition to the three CDRs seen in the case of antibodies, another region of high variability has been seen within a loop (the D-E loop) between the CDR2 and CDR3 equivalent regions of both chains. The D-E loop region has less conserved structure than in antibodies and it may form part of the TCR combining site.

Analysis of the TCR CDR lengths revealed that they show a more

restricted range than antibodies. The CDR1 in both α and β chains is not longer than 12 residues which is in sharp contrast to the much longer loops seen in CDR L1 in antibodies. The CDR2 region in TCR α chains has the same length as antibody light chains. The β chain CDR2 is 14 residues long which is the same as the minimum length for CDR H2 in antibodies. TCR CDR3 regions vary in length from 5-15 residues. This is a broader range than antibody κ chains but a much more restricted range than antibody heavy chains.

Multiple sequence alignments of TCR sequences of all possible combinations of peptide and MHC restriction was carried out by PILEUP. These sequences can be roughly divided into four groups depending on their MHC restriction and peptide specificity.

The diversity of the sequences in the four groups was analysed by a calculation of the APC and MPI values. The APC values of a sequence set are generally lower than the MPI values. Only in three instances, equal APC and MPI values are seen. No significance is attached to the equal APC and MPI values seen in the TCR set which is I-E restricted and specific for pigeon cytochrome c as there are only two sequences in that particular set. The TCR Vβ set which is DR1 restricted and specific for house dust mite also has equal APC and MPI values which may suggest a significant role for these conserved residues.

When the APC and MPI values of each of the sequence sets in the four groups were calculated it was observed that when the peptide and MHC molecule was random in a particular set the values were lower. It was found that both APC and MPI values decreased as the peptide and MHC molecule was made random in

both TCR Vα (Figure 2.12) and Vβ sequences (Figure 2.13). The sequences were much more divergent as the peptide and MHC specificity was varied.

However when the overall similarity of these sequences was calculated by PLOTSIMILARITY it was found that similarity scores of sequence sets in Group 1 were comparable to those in Group 2. The sequences in Groups 3 and 4 were generally more divergent than the sequences in Groups 1 and 2. It may be significant to note that the three sequence sets [Pigcytc/I-E (Pigeon cytochrome c specific and I-E family restricted) , Betv1/DR (BETV 1 specific and DR family restricted) and Mbp/DR (myelin basic protein specific and DR family restricted) in Group 2 which show a high degree of similarity all show MHC family specific restriction. In principle it still obeys the general rule since the variation in the MHC restriction is still very small.

In the case of the sequence set which is I-E restricted and specific for pigeon cytochrome c, the similarity scores are much higher than those seen in the group with similar peptide and MHC restriction. This may be due to the fact that this particular sequence set has only two sequences and any significance is therefore difficult to establish. Similarity scores of the CDRs in this particular set were also much higher than those found in any other sequence set.

Generally the CDRs in sequence sets with the same peptide and MHC family specific restriction had a much higher similarity score as compared to those seen in sets with divergent peptide and MHC specificity. Whether the similarity in the CDRs of sequences also reflects conservation of residues at the individual positions was also examined. The amino acids were classified into four groups

depending on their physicochemical properties. Glycine, Cysteine and Proline were left out of this scheme as they do not exhibit a common behaviour.

A simple scoring matrix was derived from the Gribskov table (Table 2.7). This matrix was then used to assign scores to each CDR position in the various sets of TCR Vα and Vβ sequences in the four groups. Group substitutions were assigned higher values as compared to substitutions across more diverse groups. The average of the amino acid scores in a particular group was then calculated.

There is a high conservation of residue character at CDR positions in the sequence sets with same peptide and MHC specificity. A sharp decrease in the CDR1 and CDR2 scores is seen in TCR Vα and Vβ sequence sets with random peptide and MHC as compared to those sets with same MHC and peptide specificity.

Even though the CDR3s of TCR Vα and Vβ sequences with same peptide and MHC specificity show higher average scores as compared to sequences with random peptide and MHC specificity, the CDR3s do not show a sharp decrease in residue similarity at corresponding positions in sequence sets across the four groups (Figures 2.18 and 2.20). The CDR3 positions in the various sequence sets vary in their scores only at the central CDR positions corresponding to the most accessible portion of the loop.

CHAPTER 3

Molecular Modelling of TCR Variable Domains

Information on the 3-D structure of TCR αβ chains is currently limited to molecular models generated from the growing database of TCR sequences and the two crystal structures of the TCR Vα and Vβ single fragments. Problems in the expression of soluble TCR in large quantity and the natural presence of carbohydrate on the molecule have hindered the elucidation of the complete structure of the TCR. Considerable effort has therefore been directed at protein engineering of secreted forms of the extracellular portions of TCRs that could be amenable to crystallographic analysis.

Recent crystallographic analysis of the β chain (Bentley *et al.*, 1995) of the 14.3.d clone (specific for a HA peptide of influenza virus A/PR/34 (H1N1) in association with the MHC Class II I-E^d molecule) and the α chain (Fields *et al.*, 1995) from the 1934.4 clone (specific for the N terminal nonapeptide of mbp in association with the MHC Class II I-A^u molecule) confirms the structural similarity between T and B cell receptors. The α and β chains are members of the Ig superfamily with the extracellular portion predicted to have a structure similar to a membrane anchored form of an Ig Fab fragment.

As in the case of Fab fragments, there is one variable (V) Ig domain and one constant (C) Ig domain in each chain. The Ig domain is composed of two antiparallel β sheets interacting face to face to form a β-sandwich structure, with a disulphide link between the two sheets ('the Ig fold') [Novotny *et al.*, 1983]. The V domain is a nine stranded β-sandwich and the C domain a seven stranded sandwich. The Vα/Vβ and Cα/Cβ domains associate into dimers which have a β-barrel structure.

It was proposed that the TCR binding site for the MHC-peptide complex is composed of loops connecting the β-strands of the Vα and Vβ domains (Davis and Bjorkman, 1988; Chothia *et al.*, 1988). These loops, three from each domain are hypervariable in sequence and are termed complementarity determining regions or CDRs 1, 2 and 3. The sequence variability of TCR CDRs 1 and 2 is lower than for antibody CDRs 1 and 2 due to a more restricted V gene pool. However there is a higher degree of variability in the CDR3 region of TCR mainly due to the larger number of J gene segments for the TCR compared to antibodies. This led to the suggestion that the CDR1 and CDR2 of both the α and β chains of the TCR interact preferentially with the less diverse α helices of the MHC molecule while CDR3 contacts the more variable peptide. Unlike antibody V domains, the TCR has a fourth hypervariable loop (Jores *et al.*, 1990) that has been proposed to be involved in the interaction of superantigens (Figure 3.1).

A number of features unique to the individual α and β chains were identified in the two crystal structures. The Cβ domain contains a large insertion

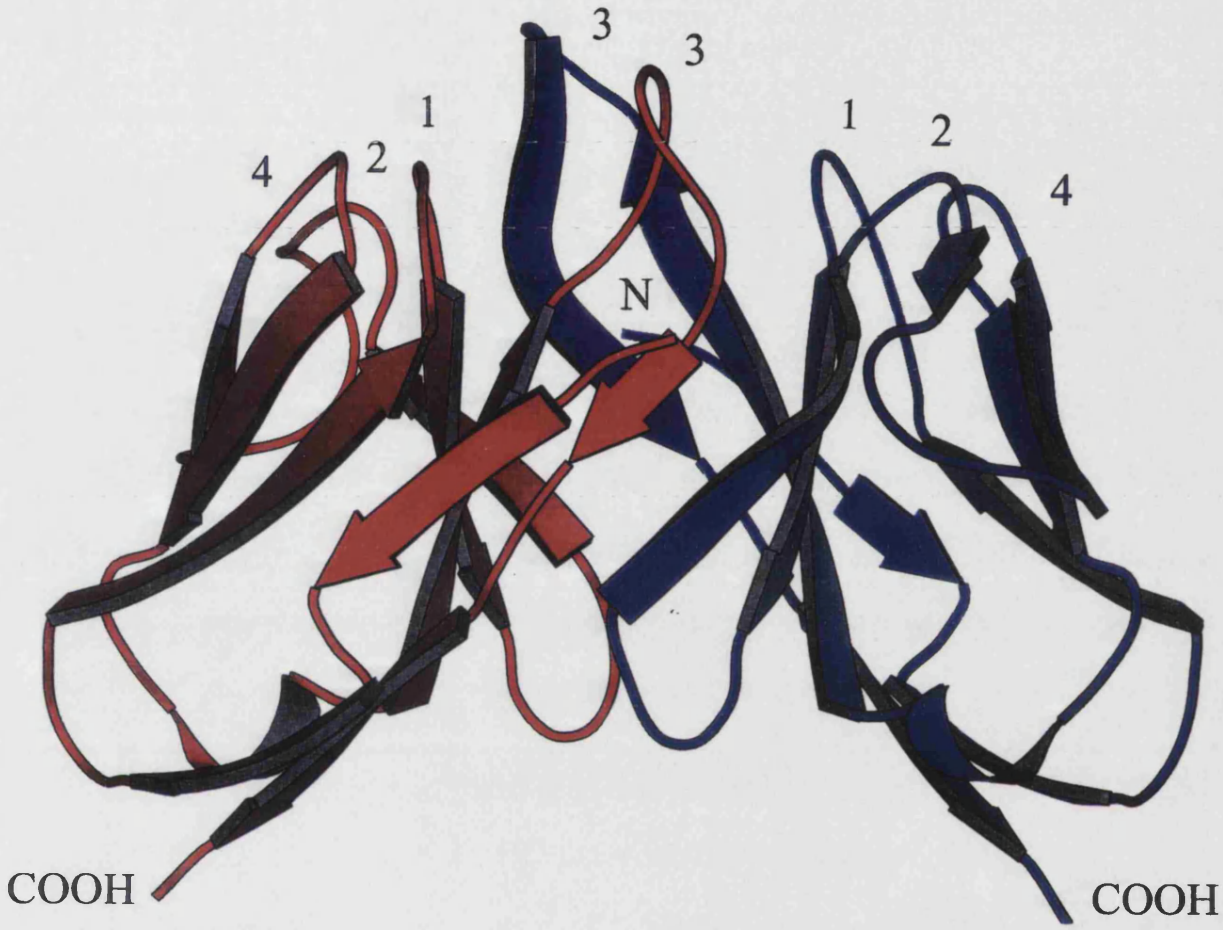


Figure 3.1 A ribbon representation of m86361 TCR VαVβ heterodimer. The CDRs of the Vα domain (red) and the Vβ domain (blue) are labelled 1-4. The figure is generated using MOLSCRIPT.

between residues 219 and 232 which forms a protruding solvent exposed loop on the external face of Vβ. Although the function of this loop is unknown its location in the 3D structure of the β chain suggests that it could be involved in contacts with the CD3 components of the TCR/CD3 complex.

The most striking feature of the Vα domain is its folding topology which differs from that of all other Ig superfamily members. There is a switch in a polypeptide strand from one β sheet to the other such that the Vα c'' strand is hydrogen bonded to the d strand of the adjacent sheet instead of the c' strand in the same β sheet as seen in antibodies and the Vβ chain. This structural difference may have functional implications. The strand switch removes a surface protusion from the molecule thereby enabling two Vα homodimers to pack as a parallel dimer of dimers in the crystal structure. Whether the αβ-TCRs dimerize in a similar manner is unknown. However a model of the $(V\alpha V\beta)_2$ tetramer was constructed from the $(V\alpha V\alpha)_2$ tetramer and known Vβ structure. This model has similarity to the MHC Class II dimer of dimers observed in HLA-DR1 crystal structures and suggests a geometry for docking the TCR with peptide/MHC complex similar to that suggested by Jorgensen et al, 1992. Another orientation of the TCR was suggested by Hong et al, 1992 which appears to be at 180° to this orientation. Whether the TCRs recognise peptide-MHC in a single/ multiple orientation is still unknown although models from this group suggest this is unlikely in some if not all complexes (Wedderburn *et al.*, 1995)[Figure 3.5].Structural information on TCR/peptide/MHC complexes is essential to resolve this issue.

In the absence of crystal structures, models for the interaction between TCR and MHC/peptide have been proposed by a number of groups including this laboratory. These models are based on the 3D structures of the MHC/peptide moiety and the expected structural similarity between the TCR and Ig Fab. Methods that were developed to model antibody V domains have been extended to generate structures for TCR V domains. In this thesis, a modification of the antibody modelling algorithm AbM called TCRM is used to generate six models of human TCR V domains, 5 molecules from DR1 alloreactive clones namely m97708, m97712, m97718, m97723, m97724 and the molecule m86361 from a DR4 restricted clone isolated from the site of inflammation in an RA patient (Table 3.1). The TCR α and β chain sequences are analysed prior to their use in the modelling method.

3.1 TCR Sequence analysis

TCR alpha and beta chain sequences were obtained via anonymous ftp from the OWL non-redundant protein database running under VAX/VMS which contains entries from several DNA and protein databanks (SWISSPROT, GENBANK, NBRF-PIR, BROOKHAVEN, PDQ-KYOTO, NEWAT86, NBRF-NEW). The sequences were then aligned pairwise and ordered as previously described. This list was then edited so as to only contain one copy of each of the variable gene segments. The aligned sequences were truncated two positions after the second conserved cysteine and then realigned using a block of antibody sequences. The antibody V_H and V_L domains were initially structurally aligned to

TCR	Clone
m86361	WM-3
m97708	AL62.24
m97712	S14.11
m97718	S14.6
m97723	S26.2
m97724	S30.10

Table 3.1: The T lymphocyte clones which express the modelled TCRs

give conservation of the forty conserved residues described by Chothia *et al.*, (1988). Structural information was included in this final alignment procedure as the β sheet framework in the antibodies was given a greater gap penalty than the loop regions. Both the alpha and beta chain sequences show higher average identity to light chains than to the heavy chains. The distribution of the identity along the sequence differed for light and heavy chains. The region making up the D-E strands is much less conserved in α and β sequences than in antibodies.Detailed analysis of the frequencies of occurence of residues at positions equivalent to positions vital for maintaining the structure of antibodies was carried out to quantify the light and heavy chain character of the two types of TCR sequences (Searle, unpublished). Positions 15, 157 and 159 which are involved in V_H-C_H contacts are also conserved in light chains. Both α and β chains showed greater homology to the heavy chains at these positions than to light chain sequences. Positions 46, 54, 56, 128 and 150 which are involved in V_L-V_H contacts show greater homology to light chains rather than heavy chains. There is virtually no heavy chain character at four of these five positions while position 54 in both α and β chains show some light and some heavy character.

To quantify the light and heavy chain character of the TCR frameworks, a new similarity index was devised to compare one set of sequences to two others (Searle, unpublished). The score for each residue is calculated as the sum of the differences between the frequencies of each residue type in the two chain types at each position. The calculation was performed for TCR α and β chain framework regions (residues 1-22, 34-48, 62-91, 105-118 in the α chain and residues 1-23,

34-47, 65-92, 107-120 in the β chain). In figures 3.2 and 3.3 where the line is above 0 the TCR sequences show more light chain character and where it is below heavy chain character.

The scores at each framework position can be summed for each chain and give an indication of the overall homology of the test chain type to the others. For TCR α chains the light chain score was 20.9 and the heavy chain score was 7.44. For TCR β chains the scores were 19.56 and 9.21 respectively. Light chain dimers were used as the framework structure in building the model as both the α and β chains appear to have greater light chain character especially in the domain packing residues.

3.2 TCR Modelling

There are two essential factors to be considered prior to modelling of TCRs. The sequences of both the α and β chains of a single clone must be known. The peptide antigen and MHC specificity should be known. Ideally the MHC restriction should be to an MHC molecule with a solved structure. Using these restrictions the choice of sequences for use in modelling becomes very limited. Only 14 pairs of human TCR sequences could be acquired from the database. Of these 14 pairs, eleven were from DR1 restricted alloreactive clones while two were DR1 restricted and specific for Influenza HA peptide 307-319. One pair was isolated from a DR4 restricted T lymphocyte clone present at the site of inflammation in a RA patient.

The fine specificity of TCR interaction with the influenza hemagglutinin

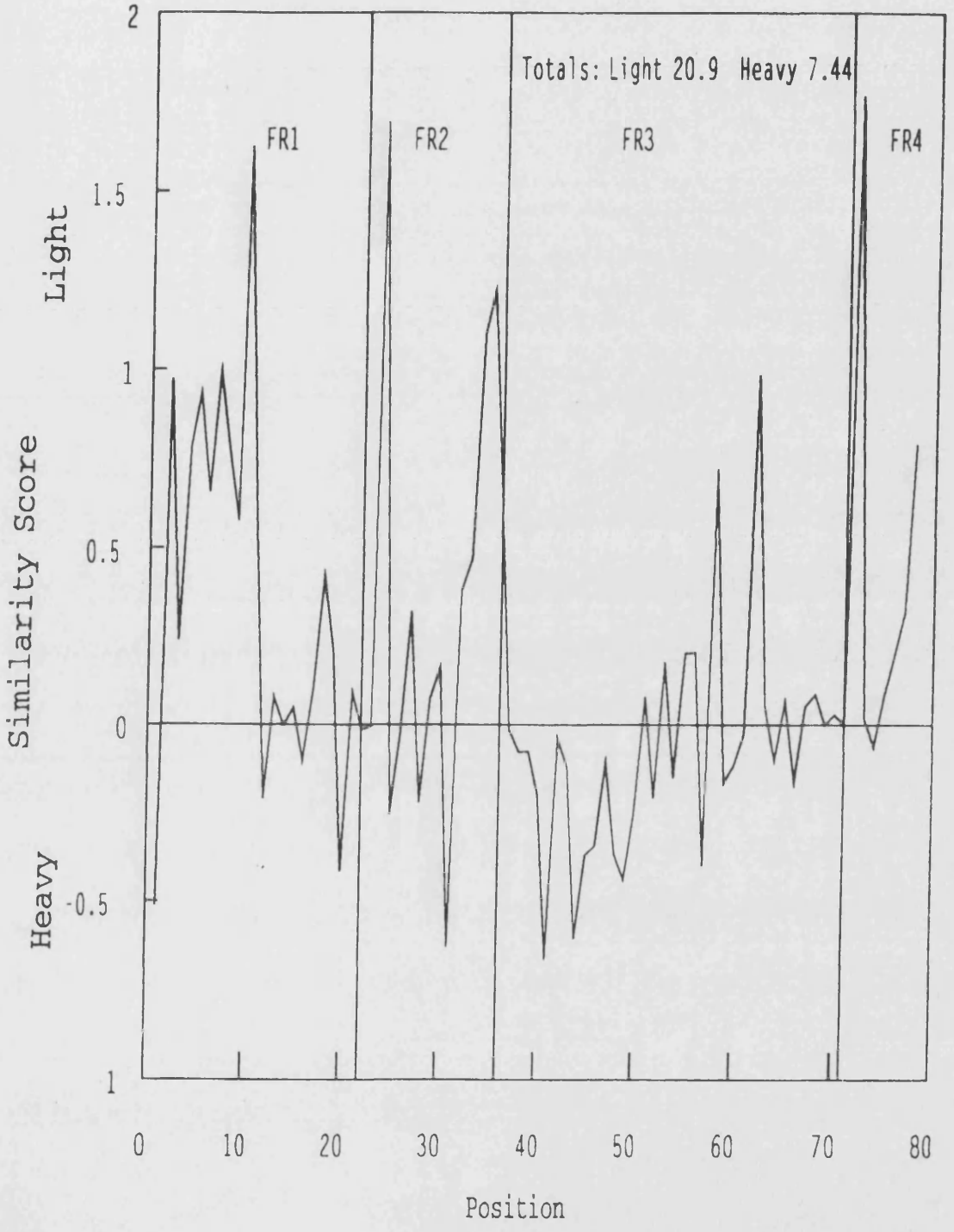


Figure 3.2 Similarity plot for TCR Vα chains. Negative values indicate greater similarity to antibody heavy chains while positive values greater similarity to antibody light chains. Only the framework regions are included in the plot. The total scores shown are the sum of positive scores and the sum of negative scores indicating the overall similarity to antibody light and heavy chains. Overall similarity of α chains is greater to light chains than to heavy chains.

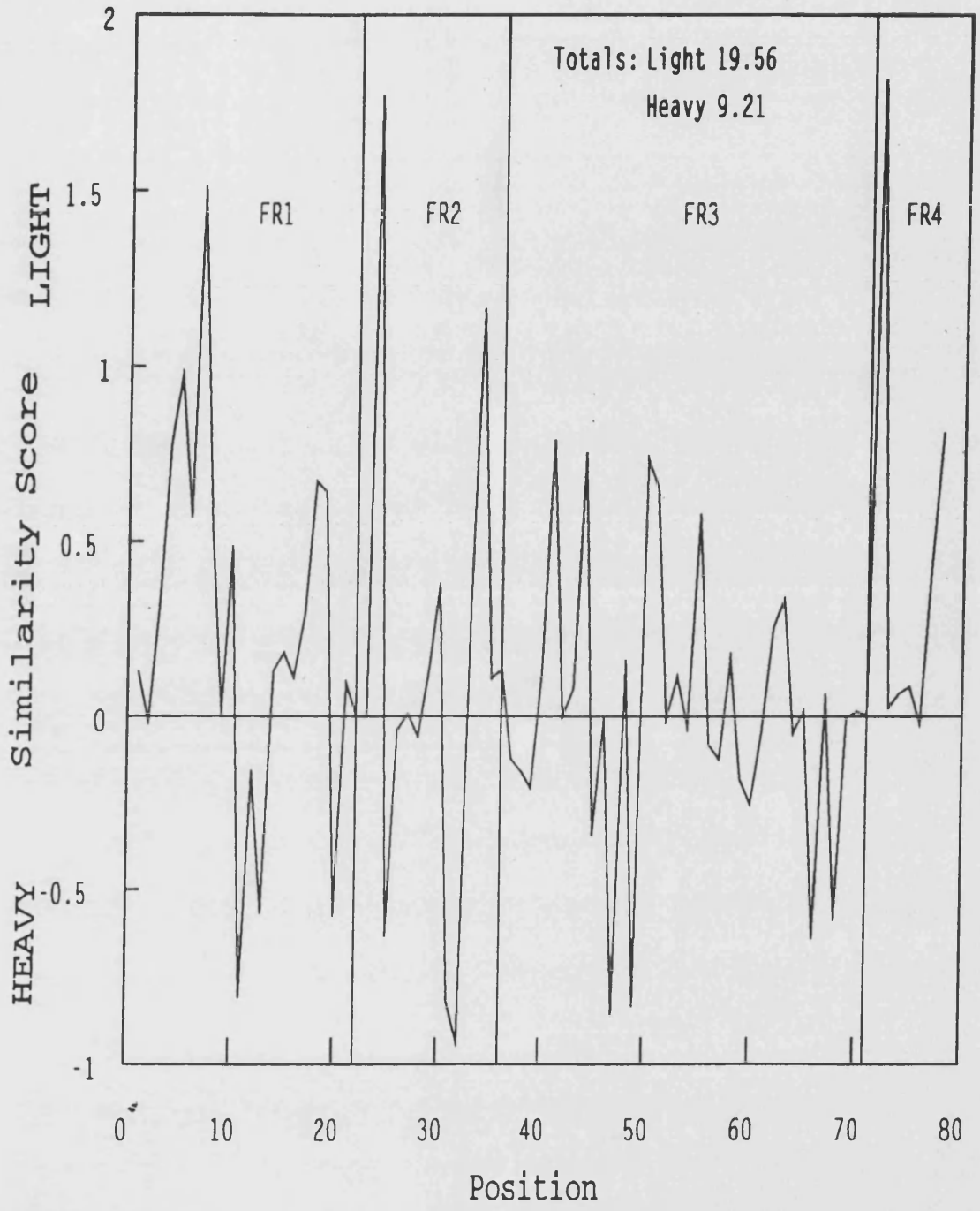
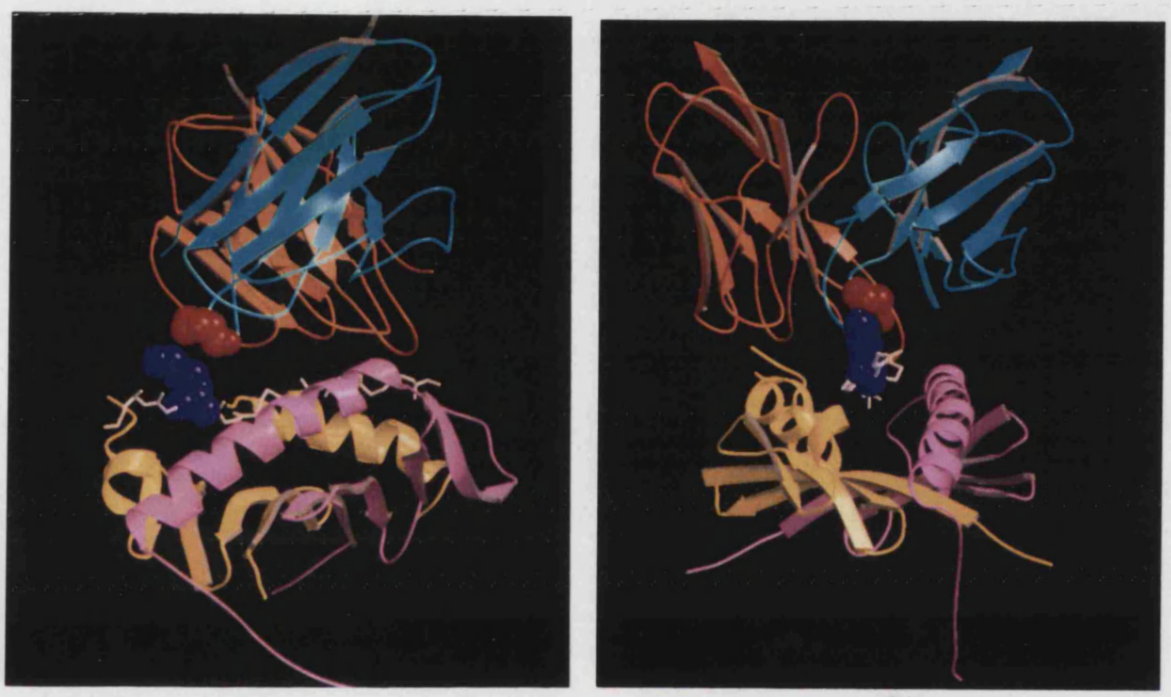


Figure 3.3 Similarity plot for TCR Vβ chains. Negative values indicate greater similarity to antibody heavy chains while positive values greater similarity to antibody light chains. Only the framework regions are included in the plot. The total scores shown are the sum of positive scores and the sum of negative scores indicating the overall similarity to antibody light and heavy chains. Overall similarity of β chains is greater to light chains than to heavy chains.

peptide HA 307-319 in the context of the DR1 (DRA, DRB1 0101) and DR4 (DRA, DRB1 0404) was studied in the two human T cell clones derived from different individuals. Models of these two TCRs (HA1.7 and cl-1) specific for the peptide HA 307-319 were generated and they showed that the conserved 94E in the TCR Vα chain could interact with the lysine at position 316 in the peptide, a known TCR contact residue. The molecular model from the TCR HA1.7 was subsequently docked on to the HA307-319/DR1 complex and the resulting complex was used to direct site specific mutagenesis experiments (Wedderburn *et al.*, 1995) [Figures 3.4 and 3.5].

Another study was carried out by Hurley *et al.*, 1993 to understand the biological consequence of DR1 microvariation. Class II molecules DR1 are naturally occuring variants found frequently in the human population. TCR utilised by human alloproliferative T lymphocyte clones which can distinguish between two microvariants of DR1 [DR(α,β1*0101) and DR(α,β1*0102)] were characterised by cDNA sequencing (Hurley *et al.*, 1993). The two DR molecules differ only at β-chain residues 85 (Val/Ala) and 86 (Gly/Val). In the DR structure these residues lie on an α- helix at the top of the DR molecule near the end of the antigen-binding cleft. Although the amino acid sequence differences appear to be subtle, DR1 molecules encoded by the two alleles may bind different repertoires of antigenic peptides and can be distinguished *in vitro* by alloreactive T lymphocytes. The effect of this limited diversity on aspects of the normal human immune response such as the binding of specific antigenic peptides and on the selection and utilisation of the TCR repertoire may be substantial. These clones were characterised by cDNA



A) B)

Figure 3.4 Molecular models showing the predicted interaction between TCR HA1.7 and the HA 307-319/DR1 complex. The Vα and Vβ chains are shown in blue and orange, respectively and the MHC α and β chains in pink and yellow, respectively. The peptide is shown in white and is orientated N terminus (right) to C terminus (left) in A) and B).
A) Sideview of the TCR- MHC complex showing the proximity of residue TCR Vα 94E (space-filled red) to peptide residue 316K. B) End view of the TCR-MHC complex obtained by rotation of view A) by 90°. The CDR3 loops of the Vα and Vβ chains can be seen to be located above the peptide while the CDR1 and CDR2 regions are positioned over the MHC helices. Figures were taken from Wedderburn, L.R *et al.*, 1995.

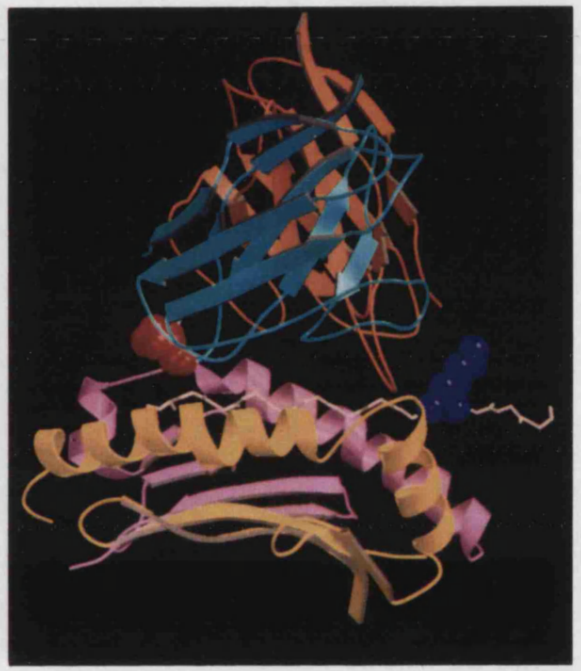


Figure 3.5 View of the TCR-MHC complex in which the peptide-MHC complex has been rotated through 180° with respect to the TCR. The peptide residue 316K is seen to be outside the interaction interface. Figure was taken from Wedderburn *et al.*, 1995.

sequencing to analyse the effect of DR1 microvariation on TCR diversity and to understand the implications of that diversity on normal immune responses and in transplantation. In this thesis, five of these TCR molecules specific for the DR ($\alpha\beta1$*0101) microvariant have been modelled. In addition to these five models the TCR from a DR4 restricted T cell clone WM-3 isolated from the synovial fluid of a RA patient was also modelled.

TCRM is a modification of the antibody modelling algorithm AbM designed to predict the structure of the TCR VαVβ heterodimer. It combines the advantages of knowledge based and *ab initio* methods. CDR loop conformations generated by a combination of *ab initio* calculations and selection of loop conformations from structural databases based on sequence homology and loop length were built on to an antibody framework (Martin *et al.*, 1989; Searle and Rees, 1996 Unpublished). This method is carried out in a series of steps which are outlined below.

3.2.1. Sequence alignment

The sequences from the six clones were loaded onto TCRM separately and the alpha and beta chains were aligned against the sequence of light chain dimers including Rei. Gaps are inserted so as to maximise the similarity of the target sequence with the database sequences. The sequence is considered to be aligned when it is the same length as the antibody sequences against which it is being aligned. The framework regions of antibodies (denoted LFR1-4 and HFR1-4) are highly conserved and are normally easy to align.

3.2.2. Framework selection

The alpha and beta target sequences are compared with each of the light chain sequences in the database in turn, giving a series of similarity scores. This similarity calculation is designed to find the database sequence which most closely matches the conserved framework regions of the target sequence. Apart from the beta chains of models m97723 and m97724, all the alpha and beta chains in the six models are most homologous to the Rei light chain. The similarity scores of each of the chains is shown in Table 3.2. TCR frameworks were modelled using the light chain dimer Rei (Epp *et al.*, 1975). Where the Rei and TCR sequences differed the TCR sidechains were inserted using a maximum overlap approach (MOP) and the framework minimised to a derivative of 0.1 in Discover (Biosym). In the MOP procedure a framework is chosen from one single structure on the basis of sequence similarity. Loops are then sampled from the Brookhaven database, which fit the required length. These loops are then scored according to sequence identity and the most similar loop is chosen as the final conformation.

3.2.3 Side chain construction

The CDRs were modelled onto this framework using CAMAL (Martin et al, 1989). The loop regions modelled in the α and the β chains in the six models are shown in Table 3.3.

For loops of less than six residues, CONGEN (Bruccoleri and Karplus, 1987) was used to generate conformations. The lowest energy conformation was chosen.

TCR	Chains	Light chain dimer	Similarity score
m86361	Vα	1rei	28.30
	Vβ	1rei	31.13
m97708	Vα	1rei	33.01
	Vβ	1rei	29.52
m97712	Vα	1rei	32.38
	Vβ	1rei	30.48
m97718	Vα	1rei	37.50
	Vβ	1rei	27.10
m97723	Vα	1rei	31.37
	Vβ	1bec	36.04
m97724	Vα	1rei	33.01
	Vβ	1bec	51.82

Table 3.2: Similarity scores of the alpha and beta chains to the most homologous light chain dimer framework

TCR	CDR	Sequence	Position	Length
m86361	A1	NYS[S(SVS)V]YLF	24-34	11
m97708		NFS[D(SVN)N]LQ	24-33	10
m97712		SYT[V(SGL)R]GLF	24-34	11
m97718		AYE[N(TAF)D]YFP	24-34	11
m97723		NFS[D(SVN)N]LQ	24-33	10
m97724		SYS[V(GIS)A]LH	24-33	10
m86361	A2	YL[S(GST)L]V	50-57	8
m97708		IPSGT	49-53	5
m97712		LYSAGEE	50-56	7
m97718		IRPDVSE	50-56	7
m97723		IPSGT	49-53	5
m97724		LSSGK	49-53	5
m86361	A3	AVRFS[L(QGG)S]EKLV	92-105	14
m97708		AVSA[S(GAG)S]YQLT	88-100	13
m97712		AVPT[T(TDS)W]GKLQ	90-102	13
m97718		AASKRG[A(GGT)S]YGKLT	91-106	16
m97723		AVG[D(YGQ)S]FV	88-97	10
m97724		AVY[S(QGA)Q]KLV	88-98	11
m86361	A4	NKSQT	68-72	5
m97708		VATER	64-68	5
m97712		TKKE	67-70	4
m97718		[N(KSA)K]	67-71	5
m97723		VATER	64-68	5
m97724		NIQEK	64-68	5
m86361	B1	SPI[S(GHK)S]VS	24-33	10
m97708		SPI[S(GHT)S]VY	24-33	10
m97712		QVD[S(QVT)M]MF	24-33	10
m97718		RSL[D(FQA)T]TMF	24-34	11
m97723		QVD[S(QVT)M]MF	24-33	10
m97724		EQNLNHDAMY	24-33	10
m86361	B2	QY[Y(EKE)E]R	48-55	8
m97708		WY[D(EGE)E]R	48-55	8
m97712		TAN[Q(GSE)A]TY	48-57	10
m97718		TS[N(EGS)K]AT	49-57	9
m97723		TAN[Q(GSE)A]TYE	48-58	11
m97724		YSHIVNDFQ	48-56	9
m86361	B3	ASSFP[G(NRG)Y]WYGYT	92-106	15
m97708		ASSL[G(QGT)Y]EQY	92-103	12
m97712		SVA[G(QGT)D]EQY	94-104	11
m97718		SAS[K(TGR)G]EQY	95-105	11
m97723		SVED[R(DRV)Y]NEQF	94-106	13
m97724		ASS[L(DSW)D]TQY	92-102	11
m86361	B4	FPNY	69-72	4
m97708		FPNY	69-72	4
m97712		PNLT	71-74	4
m97718		[H(ASL)T]	71-75	5
m97723		RPNLT	70-74	5
m97724		EKKE	69-72	4

Table 3.3 Sequence position, length and conformational search construction scheme for each of the 48 loops, [ ]=construction area, ( )=chain closure.

Knowledge based methods which rely on the crystallographic database of protein structures have had limited success in predicting loop conformations in general (Tramontano *et al.*, 1989). The MOP or canonical structure approaches have succeeded in modelling only five of the six CDRs in antibodies. This is due to the fact that the third CDR of the heavy chain, H3 is more variable in sequence, length and structure than any of the other CDRs. This extra variability arises from V-D-J-C splicing. To deal with the CDR H3 problem several groups have attempted to use *ab initio* methods to model the combining site (Bruccoleri and Karplus, 1987). In these methods the total conformational space accessible to a particular CDR is sampled. Typically the central region of the polypeptide backbone, having characteristic bond length and bond angles (rigid geometry), is constructed between the end points of the loop by a 'chain closure' algorithm. This algorithm was modified to introduce an energy minimisation procedure which greatly expanded the domain of conformational space searched during the chain closure procedure. This modification is incorporated into the conformational search program CONGEN (Bruccoleri and Karplus, 1987) which allows the user to choose any set of standard bond length and bond angles such as the CHARMM (Brooks 1983) standard geometry parameter sets. Other approaches such as minimisation or molecular dynamics either fail to saturate conformational space or are unable to deal with the problem of long CDRs. In all these methods, the selection criteria are defined in such a way so as to allow the unambiguous identification of the correct structure within the ensemble of candidates for each CDR. The position of variability in these molecules at the end of the antigen binding groove provides a unique

model of variation

For loops of six or seven residues, Cα distance constraints within the loop were used to search the Brookhaven database (Bernstein, 1977) for structural segments which matched with a tolerance of 3.5 standard deviations. Sidechains were replaced with those of the desired loop sequence in CONGEN and were energy screened using a solvent modified Gromos potential (Aqvist, 1985). A torsional screen using the initial database loops was then performed to choose the final conformation. For loops greater than seven residues a combination of database searching and conformational searching for the central five residues of the loop was performed. The energy screen and torsional screen were then performed on the generated conformations. In all cases each loop was constructed in the absence of the others. Distance constraints for antibody CDR loop regions were used to extract database conformations for the CDRs. For CDR A1 and CDR B1 the CDR L1 constraints were used. For CDR A2 and CDR B2, CDR L2 and CDR H2 constraints were used respectively. For both CDR A3 and B3 the CDR L3 constraints were used. The five residue segment predicted to be the D-E loop in both chains was also modelled using CONGEN. The final model was energy minimised in Discover to a derivative of 0.1.

Since there is no complete αβ TCR structure the accuracy of this method was tested by generating a model of CD4, a member of the Ig superfamily more distantly related to antibodies than TCR. When this model was compared with the X-ray structure of CD4, the CDR had the following backbone RMS deviations: CDRA1 and B1 (1.6 Å), CDRA2 and B2 (2.4 Å) and CDRA3 and B3 (1.3 Å)

[Wedderburn *et al.*, 1995]. This close agreement with the x-ray structure gave sufficient confidence to proceed to the construction of the TCR model.

3.3 *Comparison of the eight models generated from the DR restricted T cell clones.*

The two TCR models generated from HA307-319/DR1- specific T cell clones which also recognise peptide presented by DR4 (DRB1*0404) Class II molecules by using the TCRM algorithm (Wedderburn *et al.*, 1995) were compared to the six models generated from the DR restricted clones in this thesis.

In all these models the β sheet framework is conserved and the arrangement of the antigen binding loops is similar to that in antibodies with the third hypervariable loop of each domain sandwiched between the first and the second hypervariable loop pairs (Figure 3.6) This arrangement of CDRs has previously been used to suggest a mode of binding of the MHC to the TCR in which the first and second CDRs of each chain interact with the MHC α helices and the two CDR3 loops interact with the peptide (Davies and Bjorkman, 1988; Wedderburn *et al.*, 1995).

3.3.1 CDR1

The conformation of CDR A1 in all the eight models appears to be stabilised predominantly by hydrophobic interactions and not by intraloop hydrogen bonding as observed in CDR B1. Residues Tyr25 and Leu (at position 32 in m97724, m97708 and m97723 and at position 33 in m97712, m86361, HA1.7 and

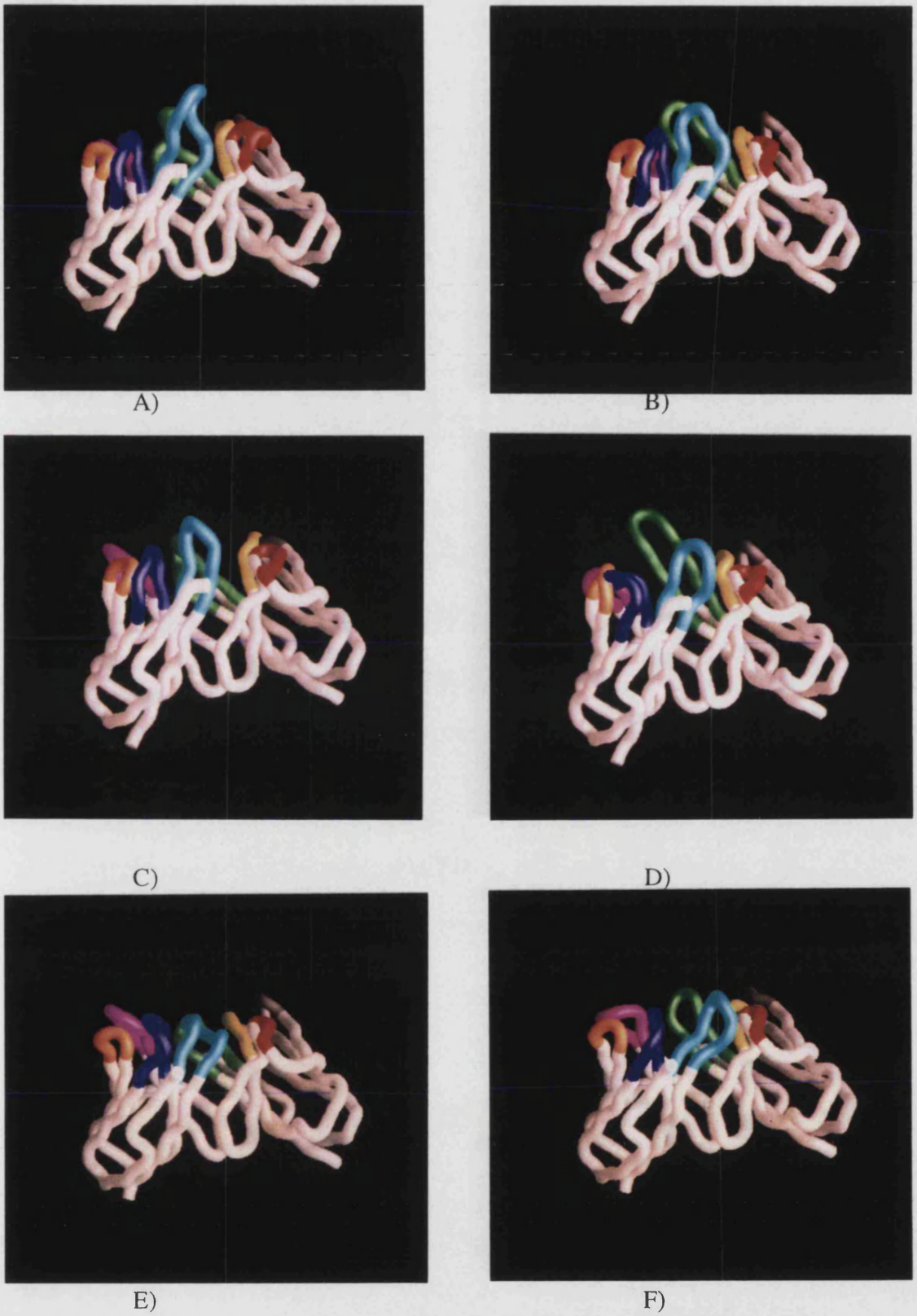


Figure 3.6 Cα backbone worms of the six models [A) m86361 B) m97708, C) m97712, D) m97718 E) m97723 and F) m97724]. Loop A1 is coloured yellow; A2, red; A3, green; A4, brown; B1, blue; B2, magenta; B3, cyan and B4, ochre. The framework regions are coloured rose pink. Figures were generated using GRASP.

cl-1) which are highly conserved in mouse and human α chain sequences pack against each other in the A1 loop. The most common substitution at both positions is phenylalanine (position 25 in m97708 and m97723 and position 33 in model m97718) which preserves the hydrophobic packing interactions (Table 3.4). In contrast, CDR1 loops in VL and VH domains is mainly stabilised by the hydrophobic side chain of residue 29 which intercalates between two β- pleated sheets. The models m86361, cl-1 and HA1.7 have identical CDR A1 which may be significant in that they contact similar α helices in the DR molecule.

The conformation of CDR B1 in the crystallised β chain is stabilised by its interaction with residue Gln25 buried within the interior of the domain beneath the hypervariable loop. Both the O and the N of the side chain amide group of Gln 25 form a total of 3 hydrogen bonds to main chain atoms of CDR B1 thus providing a supporting scaffold for the loop as it passes from one β sheet to the other. This pattern is seen in the models m97724, cl-1 and HA1.7. In the other models, position 25 is occupied by a small residue proline(2), valine(2) or serine(1). Residue 29 is located approximately midway along the Ig CDR1 forming an optimum anchorage point to stabilise the conformation of this loop. This position is occupied by Phe, Leu, Ile or Val. Only in a very small fraction a polar residue occurs at this position. In the TCR this position is mainly occupied by histidine with its sidechain partially exposed to the solvent. Histidine is substituted by valine in models m97712 and m97723 and by gln in model m97718. Either methionine or valine is present at the penultimate position (Table 3.5).

The TCRs show a more restricted range of CDR lengths than antibodies.

TCR Model	A1 loops
m97708	N **F** S D S V N N **L** Q
m97723	N **F** S D S V N N **L** Q
m97724	S **Y** S V G I S A **L** H
m97712	S **Y** T V S G L R G **L** F
m86361	N **Y** S S S V S V Y **L** F
cl-1	N **Y** S S S V S V Y **L** F
HA1.7	N **Y** S S S V S V Y **L** F
m97718	A **Y** E N T A F D Y **F** P

Table 3.4 A1 loops of the eight TCR models derived from DR1 and DR4 restricted clones. The second position is always occupied by a tyrosine or phenylalanine. Leucine is conserved at the penultimate position in all models except model m97718.

TCR Model	B1 loops
m86361	S P I S G H K S **V** S
m97708	S P I S G H T S **V** Y
m97724	E Q N L N H D A **M** Y
m97712	Q V D S Q V T M **M** F
m97723	Q V D S Q V T M **M** F
cl-1	A Q D M N H E Y **M** S
HA1.7	V Q D M D H E N **M** F
m97718	R S L D F Q A T T **M** F

Table 3.5 B1 loops of the eight TCR models derived from DR1 and DR4 restricted clones. Either methionine or valine is present at the penultimate position.

Length of A1 and B1 in TCRs is never longer than 12 residues. In these eight models, the length of A1 varies between 10 and 11 residues. In antibodies the 16 and 17 residue CDR L1 have a bulge in the side of the loop. The shorter length of the TCR CDR suggests that TCR sequences do not have a bulge in the loop.(Figure 3.7). Plot of A1 and B1 loop backbones for the six models m97708, m97712, m97718, m97723, m97724 and m86361 is shown in figures 3.8 and 3.9 respectively.

3.3.2 CDR2

There is a conserved serine at the third position in all A2 loops except in model m97718 (Table 3.6). The B2 loops in these models are varied (Figure 3.10)

Length of A2 varies between 5 and 7 which is similar to the length of Ab L2. The length of B2 varies between 8 and 11 residues which is considerably shorter than H2. The second CDRs show a variation in their length which suggests that B2 may be more involved in MHC contacts rather than the shorter A2 (Table 3.7)

As in the case of A1, the two models m86361 and cl-1 have identical A2 with a slight variation seen in the model HA1.7. It is relevant that both the A1 and A2 loops in these three models are nearly identical. This supports the hypothesis that the A1 and A2 contacts the non polymorphic residues of the DR molecule. Plot of A2 and B2 loop backbones for the six models m97708, m97712, m97718, m97723, m97724 and m86361 is shown in figures 3.11 and 3.12 respectively.

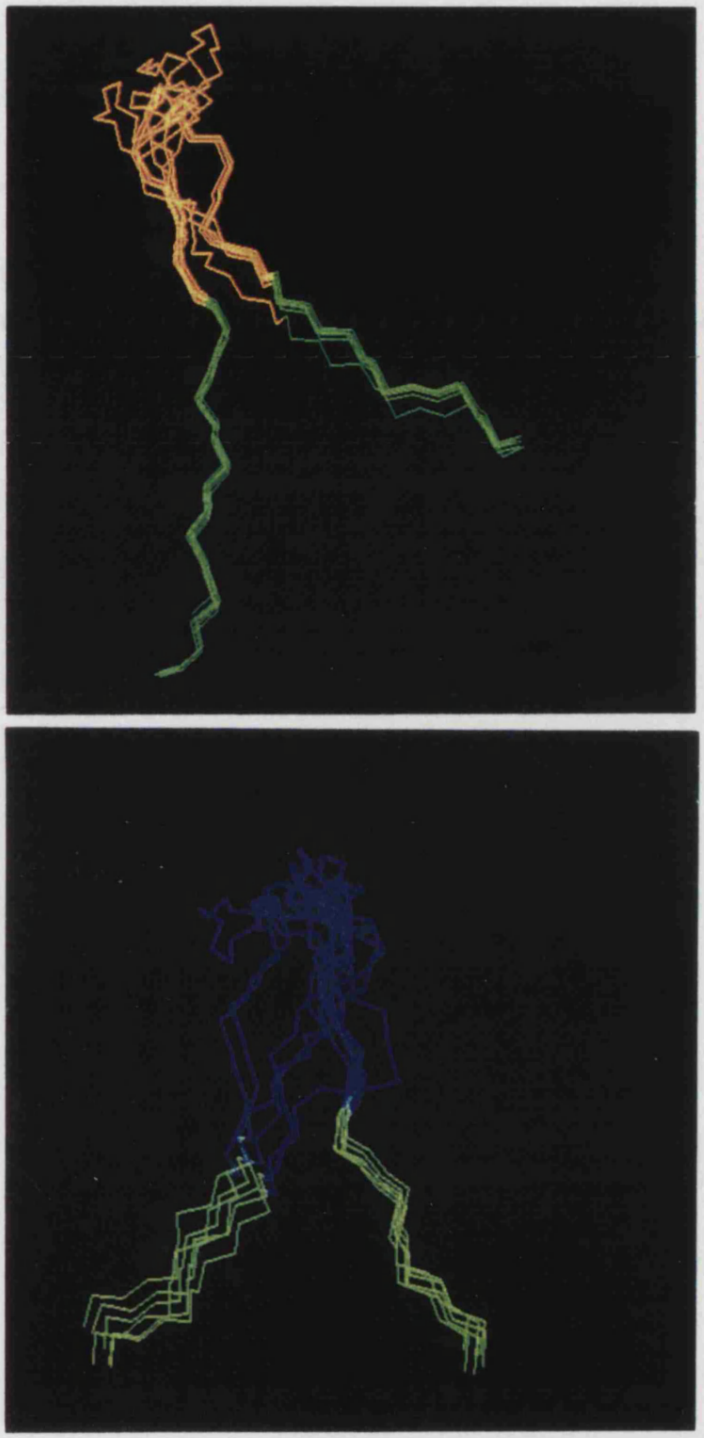


Figure 3.7 Overlapped CDR1 loops (A1 : yellow; B1 : blue) of the six models (m86361, m97708, m97712, m97718, m97723 and m97724). The framework residues are coloured green. The chains have been overlapped using five residues either side of the CDR.

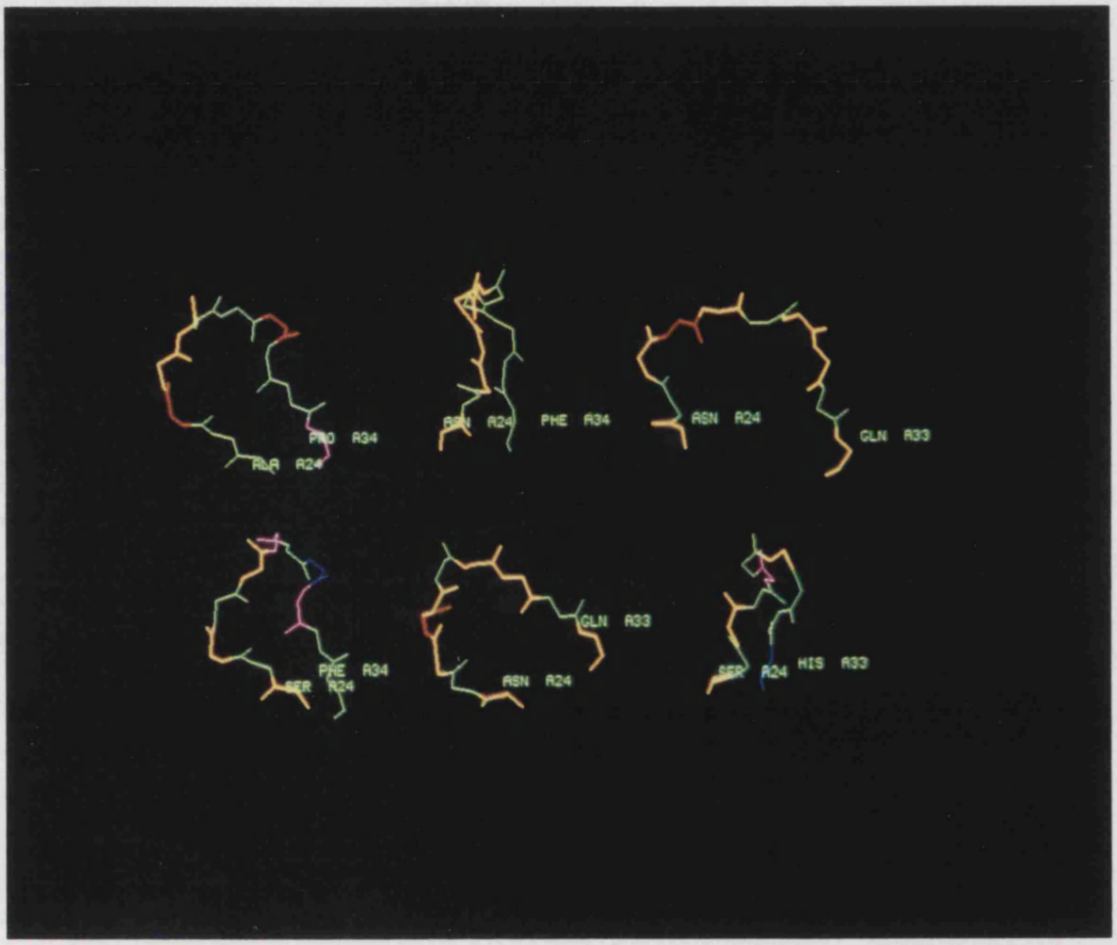


Figure 3.8 Plot of A1 loop backbones for the six models (top: m97718, m86361 and m97708; bottom: m97712, m97723 and m97724). Figures were generated using Insight.

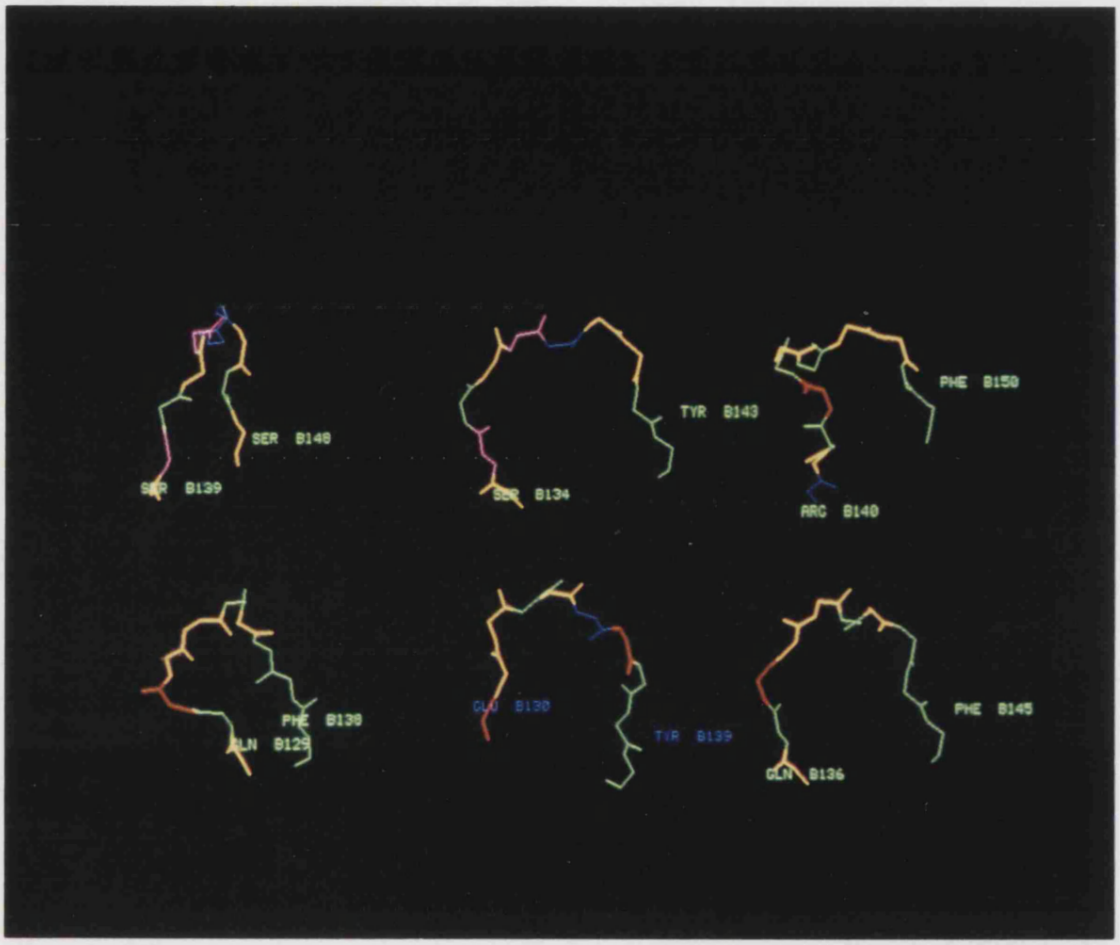


Figure 3.9 Plot of B1 loop backbones for the six models (top: m86361, m97708 and m97718; bottom: m97723, m97724 and m97712). Figures were generated using Insight.

TCR	A2 loops
m97708	I P **S** G T
m97723	I P **S** G T
m97724	L S **S** G K
m97712	L Y **S** A G E E
m97718	I R **P** D V S E
HA1.7	Y T **S** A A T L V
m86361	Y L **S** G S T L V
cl-1	Y L **S** G S T L V

Table 3.6 A2 loops of the eight TCR models derived from DR1 and DR4 restricted clones. There is a conserved serine in the third position of all models except model m97718.

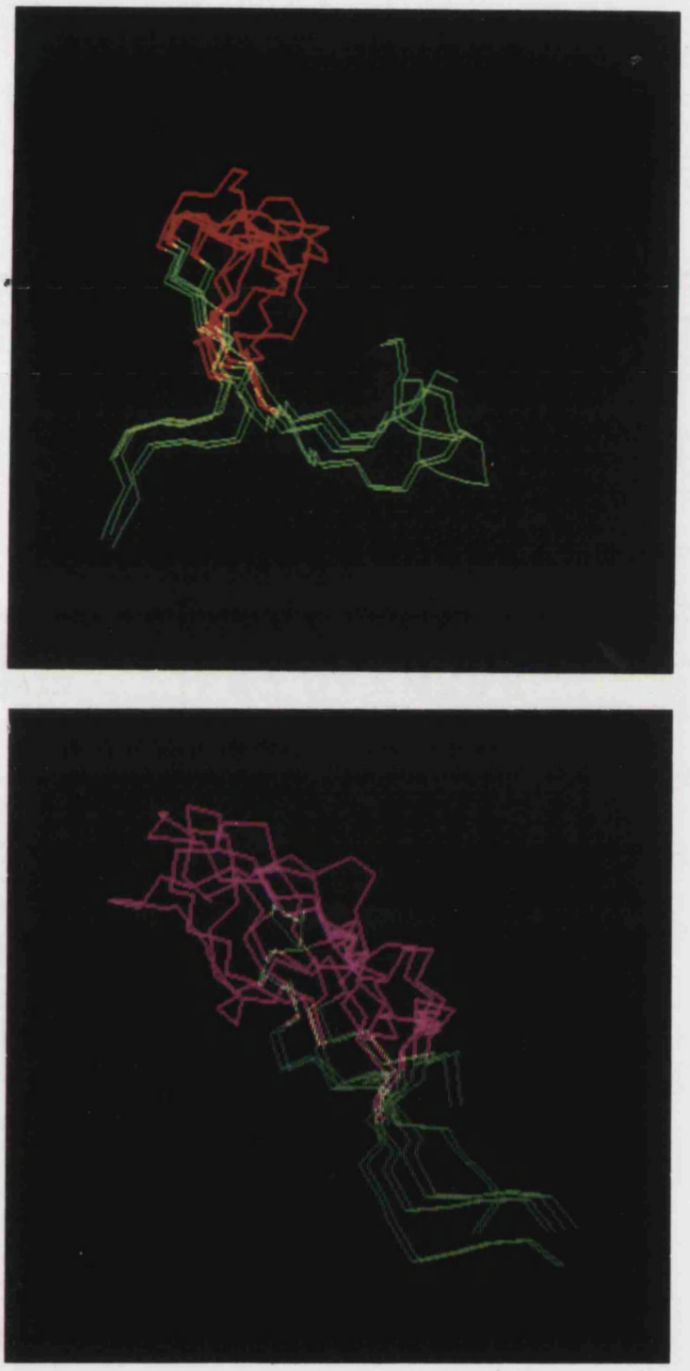


Figure 3.10 Overlapped CDR2 loops (A2: red; B2 magenta) of the six models (m86381, m97708, m97712, m97718, m97723 and m97724). The framework residues are coloured green. The chains have been overlapped using five residues either side of the CDR. Figures were generated using Insight.

TCR Model	B2 loops
m86361	Q Y Y E K E E R
m97708	W Y D E G E E R
m97724	Y S H I V N D F Q
m97718	T S N E G S K A T
m97712	T A N Q G S E A T Y
cl-1	Y S V G A G I T D Q
HA1.7	F S Y D V K M K E K
m97723	T A N Q G S E A T Y E

Table 3.7 B2 loops of the eight TCR models derived from DR1 and DR4 restricted clones.

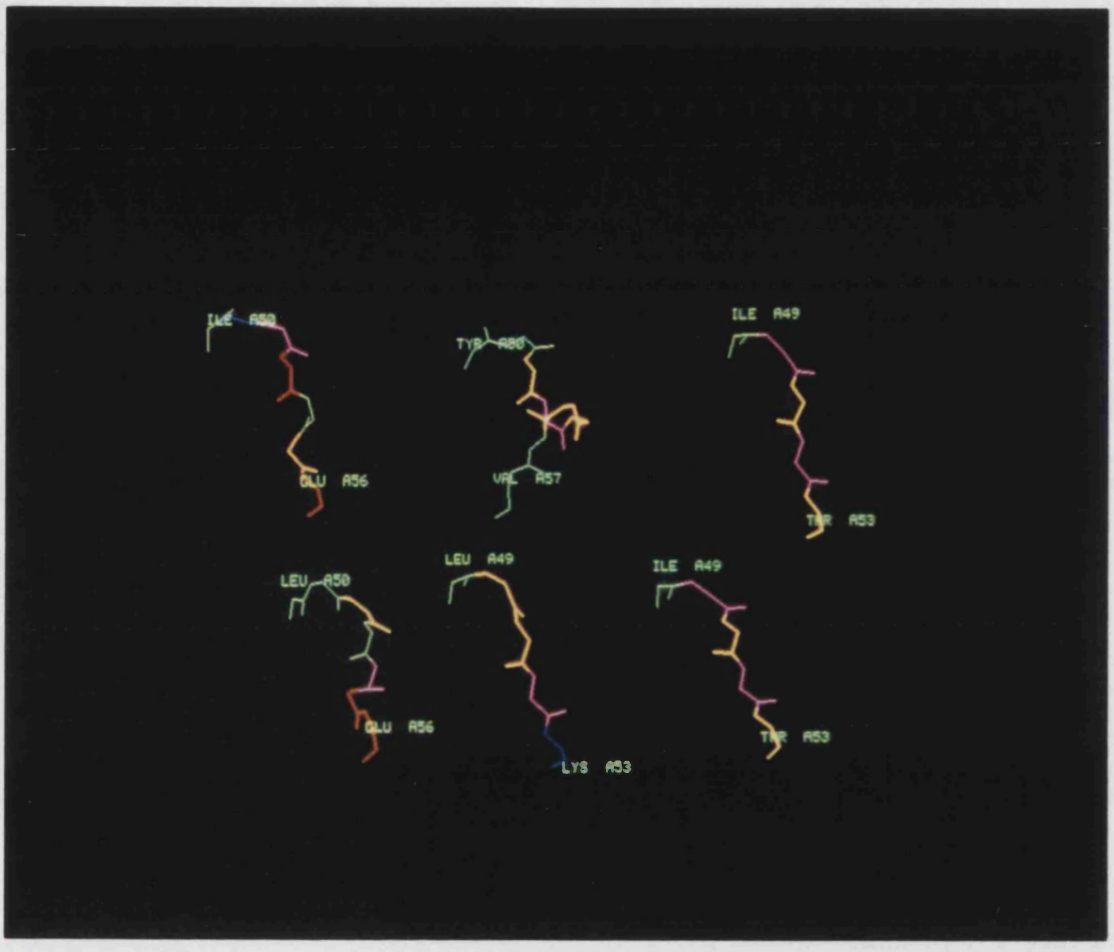


Figure 3.11 Plot of A2 loop backbones for the six models (top: m97718, m86361 and m97708; bottom: m97712, m97724 and m97723). Figures were generated using Insight.

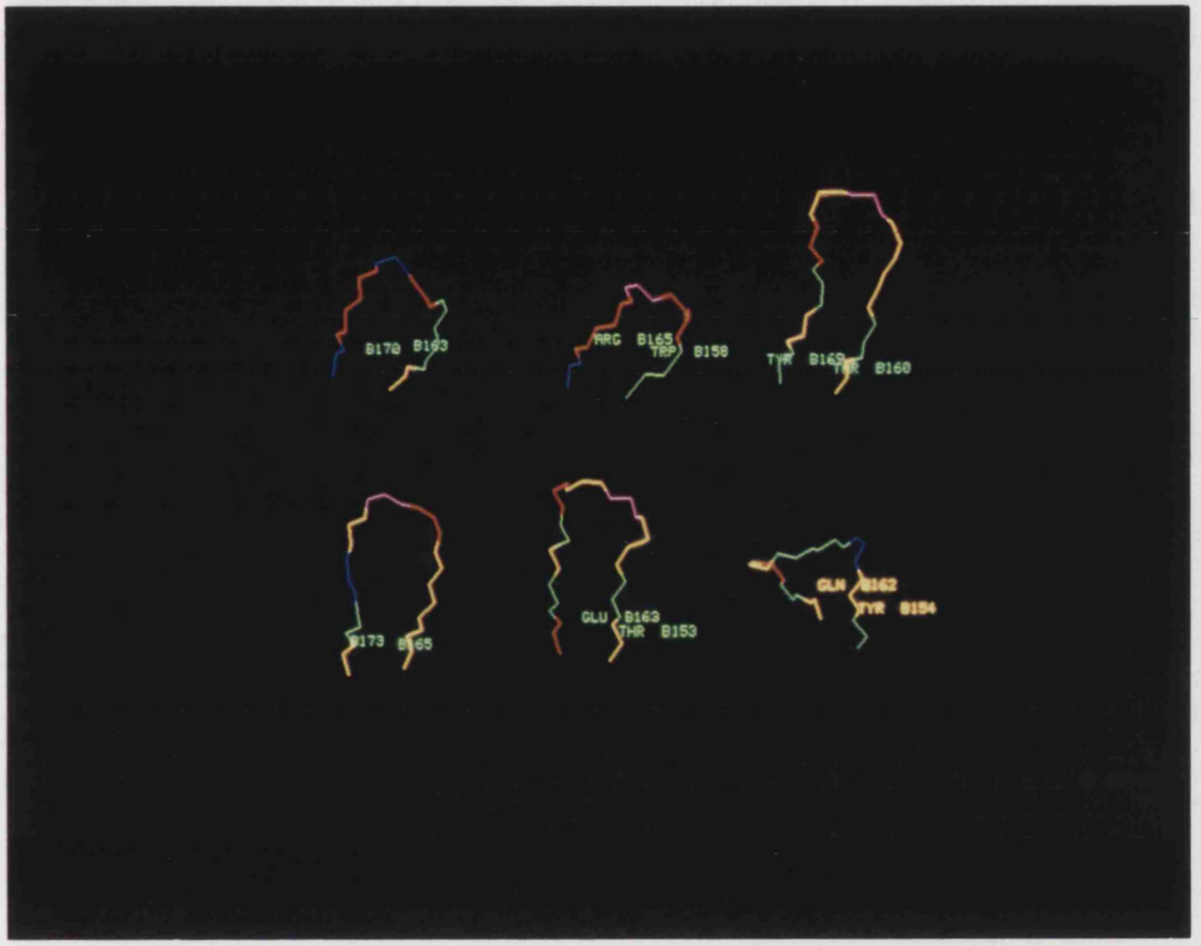


Figure 3.12 Plot of B2 loop backbones for the six models (top: m86361, m97708 and m97712; bottom: m97718, m97723 and m97724).

3.3.3 CDR3

X ray analysis of antigen antibody complexes shows that the loops created by the V(D)J junction (CDR3) of Ig H and L chains are always involved in Ag contact. Similarly the CDR3 loop of both α and β TCR chains seem critical for peptide recognition. The CDR3 length distribution of all known Ag receptor polypeptides from mouse and humans was analysed by Rock *et al* (1994).

In Ig light chains, CDR3 regions are short and of conserved length but those of H chain are long with a wide range of lengths. This may reflect the fact that Igs recognise a variety of different antigenic surfaces, from small molecules to large pathogens. For αβ TCR chains, the CDR3 length distributions are significantly constrained and are about equal in length. This probably reflects the requirement for α and β chains of the TCR to contact both the MHC and bound peptide. In γδ TCRs the γ chain CDR3 loops are short with a narrow length distribution and the δ chain CDR3s are long with a broad length distribution.

By considering all elements that contribute to the variability of the junctional (CDR3) region such as the number of D and J elements used, the D element reading frame, junctional diversity and N region nucleotide addition, Davis and Bjorkman calculated that the number of possible CDR3 sequences was the largest for γδ TCR and least for Ig (irrespective of somatic mutation) and intermediate for αβ TCR.

Length of A3 and B3 in these models varies between 10 and 16 residues. In all the models the CDRs A3 and B3 are sandwiched between the CDRs A1, A2 on one side and B1 and B2 on the other side. There is a high concentration of glycines

in both A3 and B3. The disorder that was observed in the CDR A3 loops in the homodimer was not observed in the case of these heterodimers. There is a motif AV present in all the models except model m97718 where valine is substituted by alanine (Table 3.8).

It has been proposed the CDR A3 contacts the central portion of the peptide while CDR B3 contacts the N-terminal portion. Loops A1 and A2 interact primarily with the α1 helix of the Class II molecule while B1 and B2 interact with the α2 helix.

There is a EQY motif present in the B3 loop in the 5 models derived from DR1 alloproliferative clones.(Table 3.9). The exceptions being model m97723 where tyrosine is replaced by phenylalanine and in model m97724 glutamate is replaced by threonine. This motif is not seen in models m86361, cl-1 and HA1.7. This may be significant as these three TCR molecules present peptides that are DR4 restricted. Model m86361 is characterised by a long B3 loop. Presence of common motifs in the CDR3 regions of these TCR models suggests the recognition of a limited repertoire of peptides by DR1 alloproliferative clones (Figure 3.13). Plot of A3 and B3 loop backbones for the six models m97708, m97712, m97718, m97723, m97724 and m86361 is shown in figures 3.14 and 3.15 respectively.

3.3.4 HV4

Length of A4 and B4 varies between 4 and 5 residues. The fourth hypervariable loop A4 is five residues long except in the case of model m97712 (Table 3.10). The B4 loop is four residues long except in the case of m97723 and

TCR	A3 loops
m97723	A V G D Y G Q S F V
m97724	A V Y S Q G A Q K L V
cl-1	A V S E T G G F K T I
m97708	A V S A S G A G S Y Q L T
m97712	A V P T T T D S W G K L Q
HA1.7	A V S E S P F G N E K L T
m86361	A V R F S L Q G G S E K L V
m97718	A A S K R G A G G T S Y G K L

Table 3.8 A3 loops of the eight TCR models derived from DR1 and DR4 restricted clones. The motif AV is present in all the models except in model m97718 where the valine is substituted by alanine.

TCR Model	B3 loops
m97718	S A S K T G R G **E Q Y**
m97712	S V A G Q G T D **E Q Y**
m97724	A S S L D S W D **T Q Y**
c1-1	A S R D F L S G E Q Y
HA1.7	A S S T G L P Y G Y T
m97708	A S S L G Q G T Y **E Q Y**
m97723	S V E D R D R V Y N **E Q F**
m86361	A S S F P G N R G Y W Y G Y T

Table 3.9 B3 loops of the eight TCR models derived from DR1 and DR4 restricted clones. In the five DR1 alloproliferative clones, a EQY motif preset in the TCR models. Exceptions are the model m97723 where the tyrosine is replaced by phenylalanine and model m97724 where the glutamate is replaced by threonine.

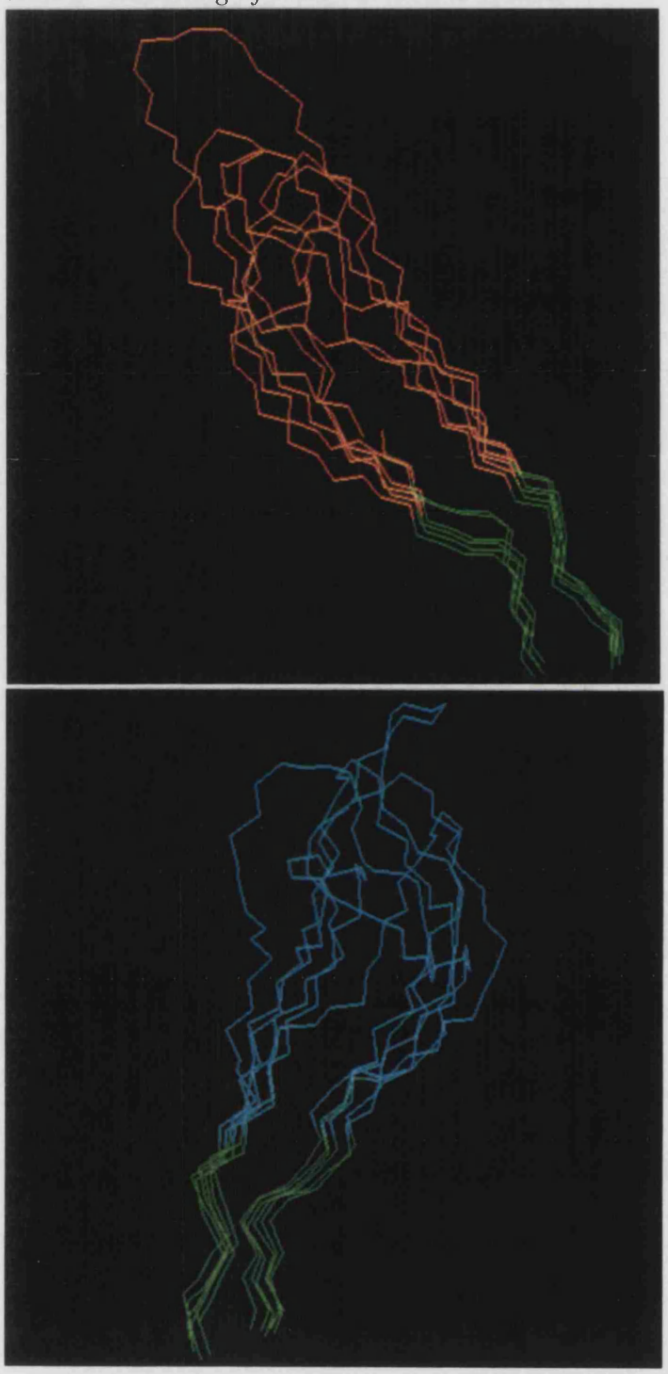


Figure 3.13 Overlapped CDR3 loops (A3: ochre; B3: cyan) of the six models (m86361, m97708, m97712, m97718, m97723 and m97724). The framework residues are coloured green. The chains have been overlapped using five residues either side of the CDR. Figures were generated using Insight.

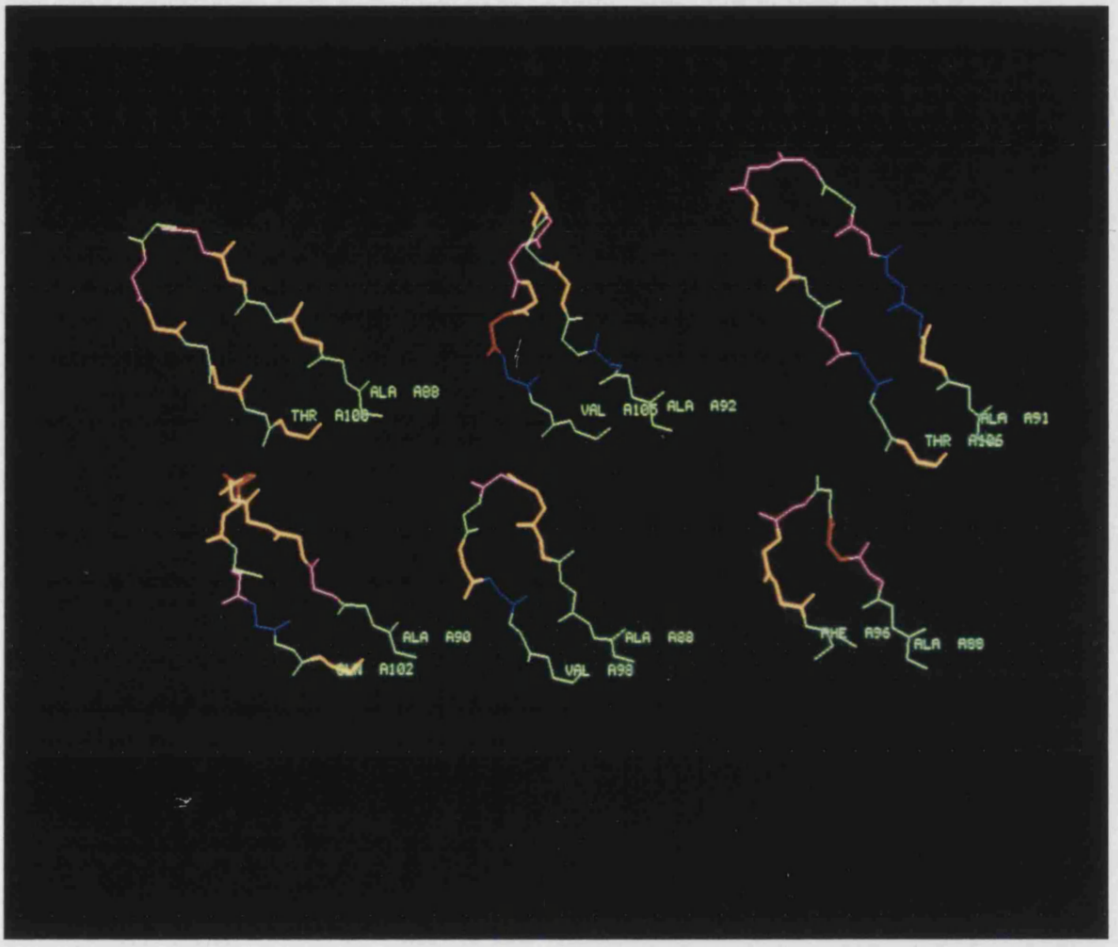


Figure 3.14 Plot of A3 loop backbones for the six models (top: m97708, m86361 and m97718; bottom: m97712, m97724 and m97723). Figures were generated using Insight.

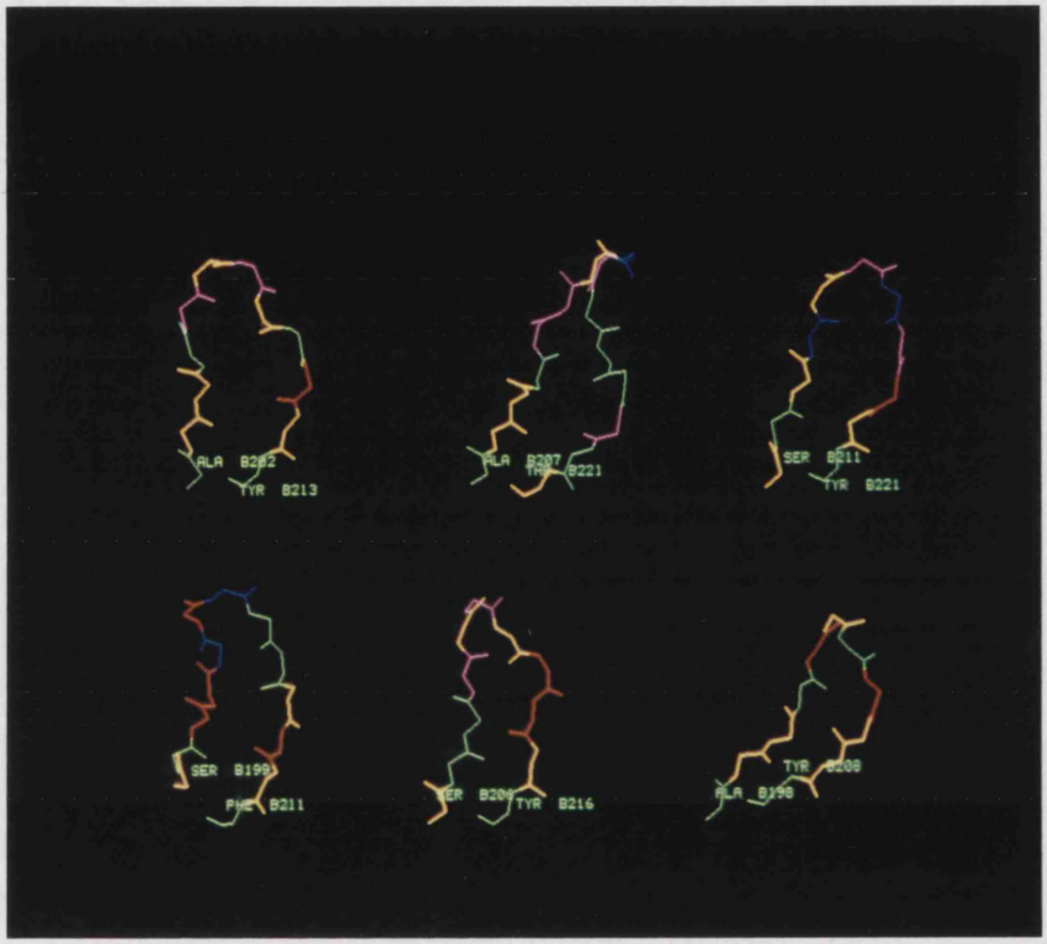


Figure 3.15 Plot of B3 loop backbones for the six models (top: m97708, m86361 and m97718; bottom: m97723, m97712 and m97724). Figures were generated using Insight.

TCR	A4 loops
m97723	V A T **E R**
m97708	V A T **E R**
m97724	N I Q **E K**
m97712	T **K K E**
m97718	N **K** S A **K**
HA1.7	**K K** S **E** T
cl-1	N **K** S Q T
m86361	N **K** S Q T

Table 3.10 A4 loops of the eight TCR models derived from DR1 and DR4 restricted clones. There is a high concentration of charged residues in all the loops.

m97718 (Table 3.11). As these loops form part of the relatively flat continuous surface with the CDRs, they could potentially be involved in interactions with the peptide-MHC complex. Plot of A4 and B4 loop backbones for the six models m97708, m97712, m97718, m97723, m97724 and m86361 is shown in figures 3.16 and 3.17 respectively.

In the A4 loops of the five models the last two residues are always charged (Table 3.10). In m86361 and cl-1, there is no charged residue seen at either of these positions.

3.3.5 Frequency of residues

In the Vα CDRs, there is a high frequency of serine, alanine, valine and glycine while in the Vβ CDRs there is a high frequency of serine, threonine, tyrosine and glutamate.

3.3.6 Speculations about canonical classes

Chothia and Lesk proposed that antibodies have only a few main chain conformations or canonical structures for each CDR. Comparison between the 3-D structures of different Fab fragments show that all the CDRs can be assigned to one of a restricted number of main chain conformations or canonical structures with the exception of the heavy chain CDR3. Sequence variations at most positions in the loops do not affect the main chain conformations but variation at a few key positions gives rise to specific interactions which cause a switch to a different canonical conformation. A canonical class is defined by the number of residues and

TCR Model	B4 loops
m86361	F P N Y
m97708	F P N Y
m97724	E K K E
HA1.7	E K K E
cl-1	S T T E
m97712	P N L T
m97723	R P N L T
m97718	H A S L T

Table 3.11 B4 loops of the eight TCR models derived from DR1 and DR4 restricted clones.

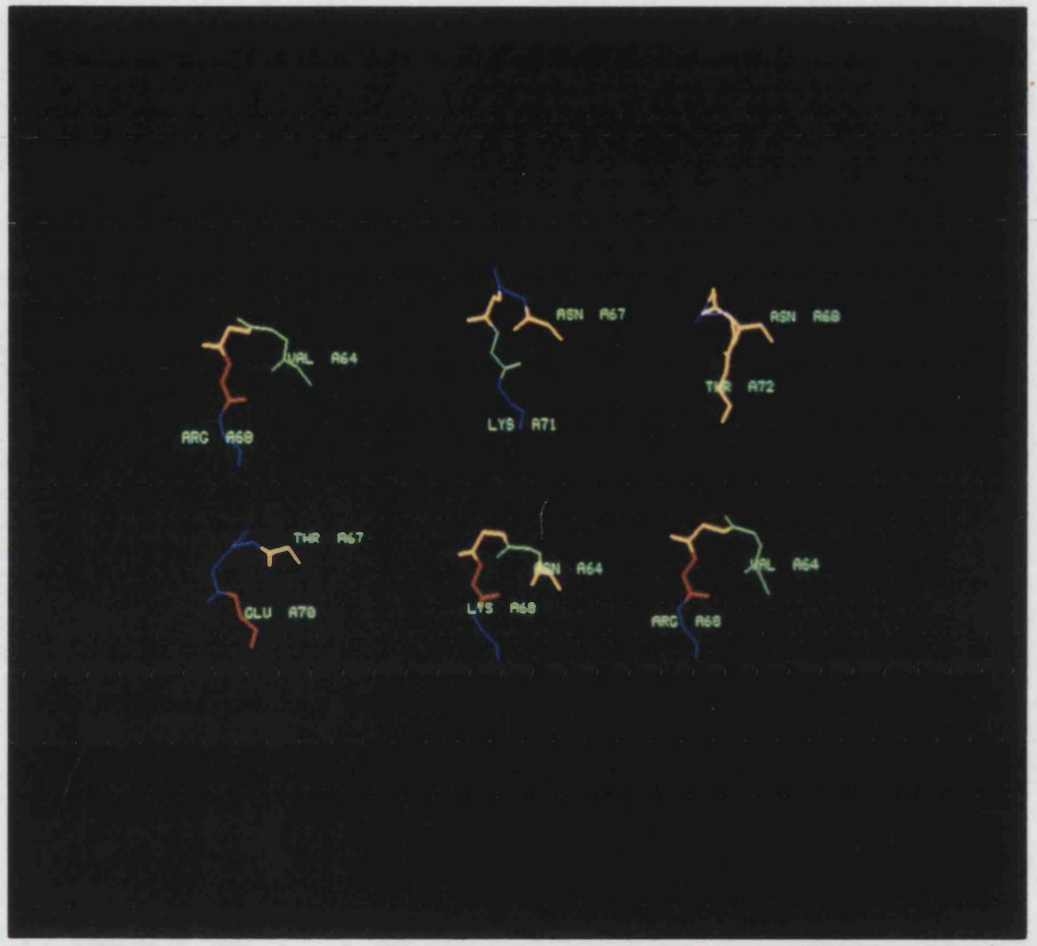


Figure 3.16 Plot of A4 loop backbones for the six models (top: m97708, m97718 and m86361; bottom: m97712, m97724 and m97723). Figures were generated using Insight.

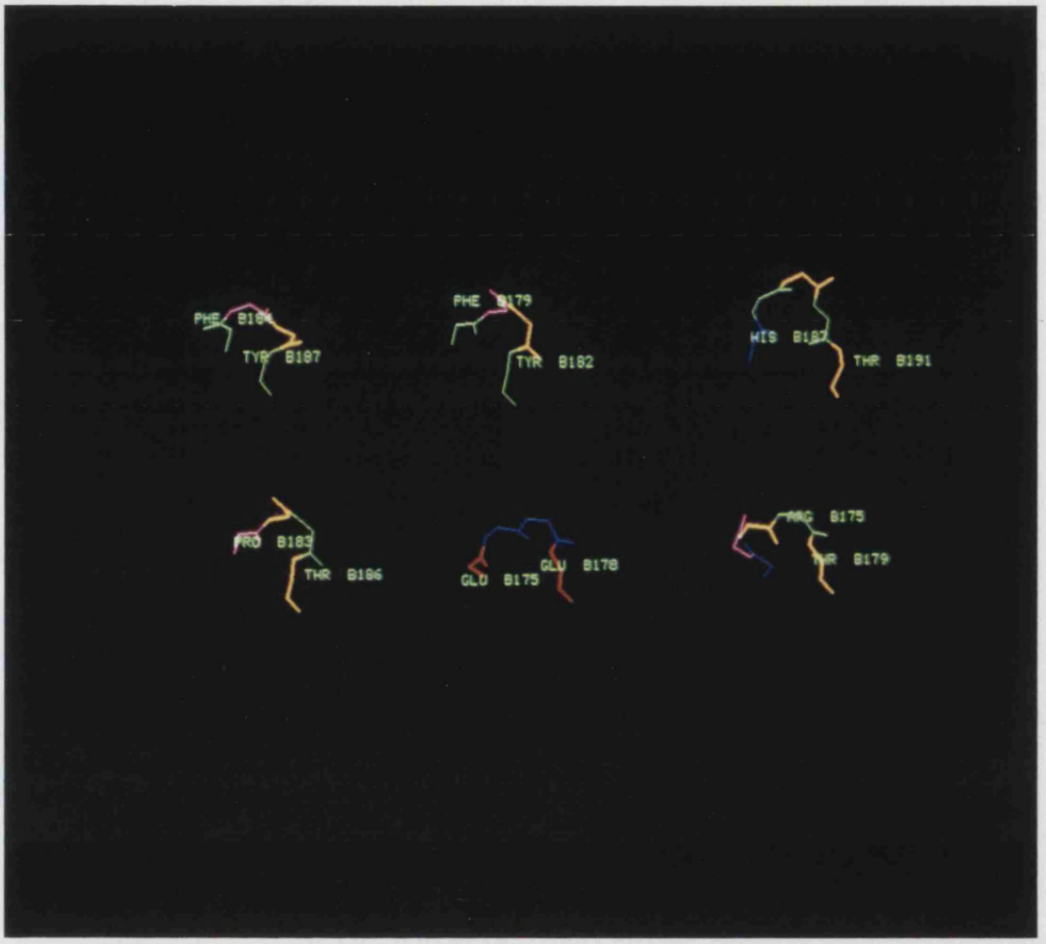


Figure 3.17 Plot of B4 loop backbones for the six models (top: m97708, m86361 and m97718; bottom: m97712, m97724 and m97723). Figures were generated using Insight.

the presence of particular residues at key positions in the loop and or surrounding framework. These classes are apparent in the large number of antibody crystal structures available. However in the case of TCR, the crystal structure of only a single Vα and a single Vβ fragment are available. The CDRs in the eight models are assorted into classes depending on the similarity in length and conformation. However these speculations can only be confirmed when several TCR crystal structures are solved.

When each CDR is examined individually, the CDRs appear to be diverse. On the basis of just differences in length and conformation, five possible canonical classes of A2 loops are seen in the TCRs from DR restricted clones namely Class1 (containing the 5 residue loops from m97708 and m97723); Class 2 (containing the 5 residue loop from m97724); Class 3 (containing the 7 residue loop from m97712); Class 4 (containing the 7 residue loop from m97718) and Class 5 (containing the 8 residue loops from m86361, HA1.7 and cl-1). Similarly 5 canonical classes of A4 loops are also seen in these TCRs, namely Class1 (containing the 5 residue loops from m97708 and m97723); Class 2 (containing the 5 residue loop from m97724); Class 3 (containing the 4 residue loop from m97712); Class 4 (containing the 5 residue loop from m97718) and Class 5 (containing the 5 residue loop from m86361).

3.4 Results and Discussion

Structural information on the TCR is limited to models mainly due to problems in producing protein suitable for crystallographic analysis. In the absence

of a crystal structure of the complete TCR, molecular modelling was carried out to predict its structure by a modification of the antibody modelling program AbM. This method makes use of the fact that the extracellular portion of the TCR resembles a Fab fragment. Prediction of the TCR structure from the primary sequence requires modelling of the conformation of the CDR loops on to the known framework or β-barrel structure.

Initially the TCR sequences were aligned against antibodies. This is relatively easy because of the high conservation of certain key residues such as the cysteines at the start of CDR1 and CDR3, a WY motif at the end of CDR1, the LLIY motif at the start of CDR2, an RF motif at the end of CDR2 and an FG motif after CDR3.

The TCR α and β chain frameworks show characteristics of both light and heavy chains. The residue motifs at the ends of the CDRs and the interface residues between the chains show more light chain character while the lengths and sequences of both CDR A3 and B3 regions are more similar to heavy chains than to light chains. The non CDR loops in the framework other than the D-E loop all have the same length as antibody κ chains. In the similarity plots, regions of the sequence were seen to be similar to each chain type but the data for interface residues suggest that the two chains may associate as in the case of a light chain dimer (Novotny and Haber, 1985). This led to the choice of a light chain dimer as the framework structure in building the TCR model. The CDRs were then modelled onto this framework using CAMAL.

Accuracy of this method was tested by generating a model of CD4, a

member of the Ig superfamily more distantly related to Abs than TCR. All the CD4 CDR equivalent regions were modelled accurately. Problems arose only in the case of CDR2 which has a different conformation to either light or heavy chains. It is more similar to a heavy chain CDR than a light chain CDR and had to be modelled using a heavy chain take off angle. There was close agreement of the CD4 model with the x-ray structure which gave enough confidence to build the model.

The six models generated by this method suggest a mode of binding of the MHC to the TCR in which the first and second CDRs of each chain were over the MHC α helices and the CDR3 loops were above the peptide. This explains the high variability of the CDR3 regions as compared to the other CDRs. CDR A1 and A2 contact one α helix while B1 and B2 contacts the other. The D-E loops interact with the MHC and are variable. The CDR A1 does not display as wide a length distribution as CDR L1 in antibodies. There are no A1 loops longer than 12 residues which means that they do not have the bulge seen in L1 loops. In the modelled arrangement of the TCR MHC complex such a bulge would clash with the MHC α helix.

Five TCR utilised by human alloproliferative T lymphocyte clones which can distinguish between two microvariants of DR1 [DR(α,β1*0101) and DR(α,β1*0102)] have been modelled. Since six different Vα and 7 Vβ segments are used by the nine DR (α,β1*0101) specific clones it appears that specific V segments are not required to recognise a microvariant of the DR1 alloantigen. The same Vα (17.1-V, 22.1, 13.1) and Vβ (2.1, 4.1-V) segments are seen in more than one TLC and these TLC are derived from different individuals. Thus it appears that

the usage of V segments in TCR which recognise DR (α,β1*0101) is non random. In the five models two Vα (13.1) and 2 Vβ (4.1-variant) gene segments are repeated. Because the majority of the primary amino acid sequence and presumably the resultant 3-D conformation are shared by the 2 DR1 molecules, use of the same V gene segments is expected. The non random V gene segment utilisation specific for DR1 molecules differs from that observed in other alloreactive TCR specific for DR4 (Vβ 6 and Vβ 8.2), DPw2 (Vβ 3.3) and B27 (Vβ 4 and Vβ subgroup 4) and has a preference for Vα 17.1-V, 22.1, and 13.1; and Vβ 2.1, 4.1, 5.1 and 18.1. The nonrandom usage may reflect the recognition of a common DR1 framework. However since the same Vβ segment can be used to recognise both Class I (HLA-B27) and Class II (DR1) alloantigens, this interaction must be flexible. In general, the number of different Vβ families utilised seem to be much more restricted than Vα families.

Since it has been postulated that the V gene segments interact with DR molecules through CDR1 and CDR2 regions these regions were examined for sequence similarities. Although shared residues and lengths are seen, these regions are found to vary even within a related subgroup. Comparison of A1 and A2 regions in these 5 DR1 models shows more extensive sharing of specific residues as compared to the B1 and B2. This conflicts with the observation noted in the original study. It is interesting to note that the three models m86361 (DR4 restricted) , cl-1 and HA1.7 (flu specific , DR1 restricted) have identical A1 loops and very similar A2 loops. It may be possible that these three TCR contact the same residues on the DR molecule. B1 and B2 loops do not show any identity. However

this difference in conservation may suggest that the α and β chains perform different functions. Majority of Vα and Vβ sequences observed are identical to previously reported sequences which suggests that the variability within the human population is limited. The two DR1 microvariants may have been selected for altered Ag binding profiles. Variation in the binding of self peptides may result in the recognition of DR1 molecules by different sets of TLC. A3 and B3 regions were then examined for common motifs which may indicate recognition of a limited peptide repertoire. Although in this particular study, a motif EQY is seen at one end of the B3 loops in the 5 DR1 models and a motif AV is seen at the other end in the A3 loops, it is not really clear what roles these motifs play in the actual recognition. It is possible that these particular five molecules bind a limited repertoire of peptides.

It can be concluded that TCRM can be used to generate models of the TCR which have predictive value. Structural information obtained from the α chain has not yet been incorporated into the TCRM algorithm. This algorithm can be improved by the incorporation of this structural data as well as any other structural information.

Chapter 4

Discussion

4.1 *Sequence analysis of TCR Vα and Vβ genes*

Various conclusions can be drawn from the sequence analysis of TCR Vα and Vβ genes. The overall diversity (calculated by determining the APC and MPI values) of different TCR Vα and Vβ sequence sets with random peptide and MHC restriction was found to be much higher than that found in TCR Vα and Vβ sequence sets with same peptide and MHC restriction (Tables 2.2 and 2.3). Typically, TCR Vα sequences from HLA-A2.1 restricted flu matrix peptide specific clones (Group 1) have APC and MPI values of 16.66 and 22.22 respectively as compared to 0.0 and 4.76 for Class I restricted sequences (with no sequenced partners) [Group 4]. In the case of Vβ sequences, the APC and MPI values of HLA-A2.1 restricted flu matrix peptide specific clones are 29.41 and 41.17 as compared to 5 and 15 for Class I restricted sequences (with sequenced partners).

The overall similarity of the sequences which have the same peptide and MHC restriction is higher than those with random peptide and MHC. Typically, TCR Vα sequences from HLA-A2.1 restricted flu matrix peptide specific clones (Group 1) have a plotsimilarity value of 0.625 as compared to 0.1 for Class I restricted sequences (with sequenced partners)[Group 4]. In the case of Vβ sequences, the plotsimilarity value of HLA-A2.1 restricted flu matrix peptide specific clones is 0.7 as compared to 0.1 for Class I restricted sequences (with

sequenced partners).

CDRs of sequences with same peptide but random MHC restriction are also more similar as compared to the CDRs in sequences with random peptide and MHC. Typically the plotsimilarity values of CDR1, CDR2 and CDR3 of DR restricted mbp specific TCR Vα sequences (Group 2) are 0.5, 0.35 and 0.325 respectively as compared to 0.05, 0.05 and 0.325 seen in the case of Class I paired TCR Vβ sequences (Group 4). Similarly the plotsimilarity values of CDR1, CDR2 and CDR3 of DR restricted mbp specific TCR Vβ sequences (Group 2) are 0.675, 0.625 and 0.625 respectively as compared to 0.05, 0.05 and 0.375 seen in the case of Class I paired TCR Vβ sequences (Group 4).

When the CDRs are individually examined by using a simple scoring scheme derived from the Gribskov table, following trends are seen.

CDR1

It is found that corresponding positions in CDR1 of sequences with same peptide but random MHC restriction (Group 2) have higher scores than sequences with random peptide and MHC (Group 4) which are less similar. Typically the average score of CDR1 in mbp specific DR restricted TCR Vα sequences (Group 2) is 0.57 as compared to 0.22 in the case of Class I restricted TCR Vα sequences (Group 4). Similarly in the case of mbp specific DR restricted TCR Vβ sequences (Group 2) the average CDR1 score is 0.55 as compared to 0.2 in the case of Class I restricted TCR Vβ sequences (Group 4).

CDR2

Similarly it was found that corresponding positions in CDR2 of sequences

with same peptide but random MHC restriction (Group 2) have higher scores than sequences with random peptide and MHC (Group 4) which are less similar.

An average score of 0.60 is seen in the case of CDR2 of mbp specific DR restricted TCR Vα sequences as compared to an average score of 0.15 in CDR2 of Class I restricted TCR Vα sequences with random peptide and MHC specificity. Similarly an average score of 0.64 is seen in the case of CDR2 of mbp specific DR restricted TCR Vβ sequences as compared to an average score of 0.22 in CDR2 of Class I restricted TCR Vβ sequences with random peptide and MHC specificity. In the case of CDR1 and CDR2, the similarity observed in the sequence sets with the same MHC and peptide specificity also reflects conservation of residue character at the individual positions.

CDR3

Even though the CDR3s of lcv specific murine TCR Vα sequences show higher average scores as compared to sequences with random peptide and MHC specificity (Class I restricted sequences), the CDR3s do not show a sharp decrease in residue similarity at corresponding positions in sequence sets across the four groups. The average scores of CDR3s do not seem to differ dramatically when the specificity is changed. In the case of CDR3, the similarity observed in the sequence sets with the same MHC and peptide specificity reflects conservation of residue character at the individual positions. The average similarity scores of CDR3 positions in sequences with random MHC and same peptide specificity (0.39 for mbp specific DR restricted TCR Vα sequences) are comparable to the scores seen in sequences with random peptide and MHC specificity (0.38 for Class I restricted

TCR Vα sequences). CDR3 positions in mbp specific DR restricted TCR Vβ sequences and Class I restricted TCR Vβ sequences vary in their scores only at middle positions as compared to average scores in other cases. Thus the nature of conservation among these sets can be studied by using a scoring scheme derived from the Gribskov table.

4.2 *TCR Modelling*

The use of new ab initio methods and the available TCR α and β chain structure enables the prediction of structures for which only the amino acid sequence is known. Using a combined algorithm which generates TCR Fv frameworks from a crystallographic database, and predicts CDR conformations by a number of methods, models with predictive value can be generated.

The residue motifs at the ends of the CDRs and the interface residues between the two TCR Vα and Vβ chains show a greater similarity to the antibody light chains. Similarity plots also suggest a high light chain character in both the chains which led to the choice of light chain dimer framework for the modelling of TCRs.

Various conclusions can be drawn from the molecular modelling of the TCR variable domains of DR restricted clones. TCR Model m86361 was generated from a DR4 restricted T lymphocyte clone isolated from the site of inflammation of a rheumatoid arthritic patient (Korthauer, U; 1992). The five other models were generated from alloproliferative T lymphocyte clones (Hurley, C.K., 1993).

TCR models show a conserved 'β barrel' structure or framework region. The characteristic folding pattern of Ig domains, first seen in the Fab structure determined by Poljak and colleagues. It consisted of two antiparallel β-sheets packed to form a β sheet 'sandwich', the two halves of the sandwich being held together by a disulphide bond.

Each CDR is a loop connecting two β- strands and has a fixed orientation on the framework depending on its length and other characteristic sequence motifs. Due to their close proximity to one another, mutational changes introduced into a CDR can be propagated to adjacent CDRs or to the framework itself. Although the three dimensional structure of Fv regions is well conserved they are not identical in all TCRs.

Depending on the overall topography of the antigen combining site, antibodies are classified into three groups, namely cavity antibodies, groove antibodies and planar antibodies. Similarly the TCR models can be differentiated on the basis of surface topology exhibited by their combining site. Models m97723 and m97724 clearly show a flat or planar topology as seen in Hy-Hel-10 antibody (Padlan *et al.*, 1989). However the four other models show a protrusion of the CDR3 loops not seen in the case of antibodies (Figure 3.6). They exist due to the difference in the binding mechanism of both antibodies and TCRs. The CDR3 loops must protrude from the surface to make contact with the peptide lying in groove of MHC helices. This suggests a reason why the TCR does not bind isolated peptide because an exposed pair of loops cannot form a strong association with the peptide. Only when the peptide lies in the MHC groove a strong non covalent association

could be formed. When the groove is empty the binding affinity of the TCR for the MHC will be reduced because water will be present in the groove reducing the hydrophobic interactions between MHC and TCR. Most of the binding energy in antibody -antigen complexes is contributed by these interactions while the specificity is produced by hydrogen bonds or charge charge interactions.

Groove or cavity type topology is not seen in these models which does not necessarily mean that they do not exist in nature. It is interesting to note that the β chain in the two models m97723 and m97724 were modelled on the 1bec light chain dimer while all the other chains in the six models were found to have greater homology to 1rei light chain dimer (Table 3.2).

The models generated so far indicate a mode of binding of the TCR to the MHC and peptide. Several features of TCRs can be explained by this arrangement. TCRs show less defined regions of variability than antibodies. In the five models, two Vα and two Vβ gene segments are used. Use of the same Vα segments is expected at the sequence level as majority of the same amino acid sequence and the resulting 3-D conformation are shared by the two DR1 microvariants. Comparison of A1 and A2 loops in these 5 DR1 models shows that there is more extensive sharing of specific residues as compared to B1 and B2 (B1 and B2 loops do not show much similarity in sequence). However when the A2 and B2 loops are overlapped it is found that they do not show a similar conformation. Identical A1 and very similar A2 loops are observed in the three models m86361 (DR4 restricted) and the flu specific DR1 restricted TCR cl-1 and HA1.7. This may suggest that these TCR molecules contact the same residues on the DR molecule.

As A1 and A2 are found to be more homologous than B1 and B2 it can be concluded that these TCR molecules will also recognise peptide/MHC complexes in a similar orientation to that seen in the case of flu peptide HA 307-319/DR1/ HA1.7 TCR complex (Figure 3.4) where the relatively conserved TCR Vα chain is placed over the invariant DRα helix and the TCR Vβ is placed over the polymorphic DRβ chain.

This suggests that TCR Vα and Vβ chains perform different functions in recognition. Generally the majority of Vα and Vβ sequences observed are identical to previously reported sequences suggesting that variability within the human population will be limited. This observation supports other studies of V gene segment diversity which have demonstrated only subtle allelic variation. When the CDR3 regions, the region of the TCR postulated to interact with peptide, was examined common motifs were found which seem to suggest that they probably bind a limited repertoire of peptides and it may suggest a similar kind of binding mechanism to the peptide antigen. A single model of TCR (m86361) infiltrating the rheumatoid synovium is also presented. This model shows also shows similarity to the TCR molecules from DR1 restricted flu HA specific clones.

4.3 ***Implications from sequence analysis and modelling***

The advent of PCR-based methods for the evaluation of TCR repertoire has increased the sensitivity and scope for investigation of T cell populations in the rheumatic diseases. PCR analysis of TCR usage in synovial samples of patients with autoimmune diseases like rheumatoid arthritis gives variable results with some

groups claiming evidence for the effect of uncharacterised superantigens expanding or deleting T cells with particular Vβ regions while others have seen oligoclonal expansion of TCR Vα and Vβ genes where a specific TCR is expanded against a background of polyclonal nonexpanded receptors.

The sequence and structural data here suggest that this simple model whereby T cell oligoclonality is dictated by the degree of variation in MHC alleles may be too simplistic. If true, it would be expected that closely similar MHC alleles containing identical peptides would elicit closely similar TCRs. This is clearly not the case. Small variations in sequence in either the α1 or α2 helix at positions that contact peptide may cause sufficiently large conformational perturbations as to reduce binding below the required threshold. However a different TCR may see this altered complex more readily. This leads to the proposal that extensive variation in T cell clonality may occur even for closely similar MHC-peptide complexes. Different MHCs are expected to select different CDR1 or CDR2 sequences or structures (different TCRs). Small changes in peptide sequence propagate changes in α1 and α2 conformation which is sufficient to disrupt TCR binding. This has been seen in the case of antibodies where a single residue difference between hen egg white lysozyme (HEL) [Q121 to H121] and turkey egg white lysozyme (TEL) causes a large drop in affinity of the murine monoclonal anti- turkey egg lysozyme antibody D1.3 (Braden *et al.*, 1996). This suggests that very small changes at a protein- protein interface can have dramatic effects on binding affinity. This leads to the proposal that extensive variation in T cell clonality may occur even for closely

similar MHC-peptide complexes. When peptides are presented in the context of different MHCs, the expected selection of different TCRs is seen.

One consequence of this effect in autoimmune diseases such as rheumatoid arthritis (RA) where an unknown antigen driven response can give rise to a large number of T cell clones. One suggestion for therapeutic intervention in RA and related diseases is the use of monoclonal antibodies to delete particular members of a polyclonal cell population (T cells, B cells etc).

Monoclonal antibodies exert their influence on the immune system in two ways namely, immunosuppression and tolerance induction. They have been effective in controlling several models of autoimmune diseases including experimental autoimmune encephalomyelitis (EAE), experimental myasthenia gravis, autoimmune arthritis and adjuvant or collagen arthritis.

In several rodent EAE models of multiple sclerosis, a limited range of T cell receptors specific for myelin basic protein-MHC class II complexes are expressed by pathogenic T cells capable of transferring disease. In the mouse PL/J strain EAE model the disease is induced by CD4 T cell clones specific for myelin basic protein peptide (MBP1-9) restricted by $I\text{-}A^{u}$ utilise a limited set of TCR rearrangements comprising a single Vα4 member with bias towards Vβ8.2 and considerable junctional conservation (Heber-Katz et al., 1989). EAE has been treated by therapies based on structural features of the trimolecular complex such as peptide analogues, anti-TCR Vβ antibodies and TCR peptides from CDR2 and CDR3 (Acha-Orbea et al., 1988; Wraith et al., 1989; Offner et al., 1991).

However in the case of RA, study of the TCR repertoire in rheumatoid synovitis gives conflicting results (Sottini et al., 1991; Paliard et al., 1991). Since no consistent pattern has emerged with respect to synovial over or under representation of particular variable regions and expansion of particular variable regions and expansion of particular junctional regions it might be impractical to target T cells directly.

Therapeutic strategies may be better targeted at the peptide binding site within the MHC where drugs may be derived that can bind to those parts of the pocket not normally susceptible to polymorphism.

APPENDIX-1

Gloop2 Fv Expression And Purification

A1.1 Introduction

A1.1.1 Structure of the antibody molecule

The antibodies form a central part of the body's defence against infectious agents and other potentially harmful foreign macromolecules. Their function is to bind to surfaces on the antigen in a high affinity complex and trigger its neutralisation and elimination, either by stimulating phagocytic cells or by activating the complement cascade to cause lysis of cells. The antibody structure combines specific epitope recognition and common effector functions in a single multidomain molecule. There are five major classes of Ig (IgA, IgD, IgE, IgG and IgM) that vary both in structural and functional properties.

An IgG molecule is a multichain protein, consisting of two identical heavy (γ-type) and two identical light (both either κ or -λ type) chains linked by disulphide bonds to form a Y shaped molecule. There are two antigen binding sites, one on each of the Fab ("antigen binding fragment") arms, and one Fc ("crystallisable fragment") region which is the site for the common effector functions. The Fabs are linked to the Fc region by the hinge region, which varies in length and flexibility in the different antibody classes and isotypes.

The light chain is composed of one variable (V) and one constant (C) domain and the heavy chain one V and three C domains. The V and C domains have

little sequence homology to each other but there is extensive homology within the V and C domains. The cysteines of the disulphide bridge between the sheets and the tryptophan that packs against it are the only residues that are conserved in all Ig domains (Lesk and Chothia, 1982). These residues form the central core of the domain interface.

Each V domain contains three hypervariable regions of sequence, termed complementarity determining regions (CDRs) (Wu and Kabat, 1970) which form the antigen combining site (ACS), while the remaining sequences in the V domain are termed the framework regions (Kabat *et al.*, 1977; Kabat *et al.*, 1987).

The immunoglobulins from the other four classes have essentially the same structure but vary in the number of C domains.

A1.1.1.1 The Ig domain structure

Early studies showed that all antibody domains form compact globular structures with a characteristic fold called the Ig fold (Poljak, 1973). Each domain is composed of 2 anti parallel sheets which form a bilayer structure stabilised by a disulphide bond between the two layers (the β- sandwich). The V domain is a nine stranded β- sandwich and the C domain a seven stranded β- sandwich. The domains (V_H/V_L, C_H/C_L) associate into dimers to form a β- barrel structure.

The domain dimers are linked by flexible polypeptide segments and there is limited interaction between domains within each chain. The hinge region between C_H1 and C_H2 is the most flexible area of the molecule and is rich in prolines and

cysteines. The length of the hinge varies in the human IgG subclasses and is related to the conformation of Ig and with activity in effector functions (Alzari *et al.*, 1988). Another site of flexibility is the switch region which links the V and C domains. A ball and socket model has been proposed for this region to describe the observed flexibility between the domains (Lesk and Chothia, 1988).

A1.1.2 The Antigen Combining Site (ACS).

The antigen combining site (ACS) is formed by the juxtaposition of six hypervariable loops or CDRs, three from the light chain V domain and three from the heavy chain V domain. Analysis of CDR conformation showed that for five of the loops, conformations were limited to a small number of structural classes called canonical structures.

Each canonical structure describes a preferred conformation for a CDR, the preference being determined by key residues whose presence controls a particular main chain conformation (Chothia and Lesk, 1987; Chothia *et al.*, 1989). Two CDRs (L1 and H3) have been proposed to be the most important in determining ACS topology and are the most variable in length (de la Paz *et al.*, 1986). H3 has the greatest heterogeneity in both length and sequence and lies in the centre of the combining site. It has a distinct role in determining specificity. Unlike the five other CDRs which have one or a small set of main chain conformations (Chothia and Lesk, 1987; Chothia *et al.*, 1989) H3 has no canonical structures. This extensive variation of CDRH3 is due to well known processes of combining D and J minigenes with possible addition of N or P

segments.

The CDRs of each chain are supported on a framework region which consists of conserved strands that fold to form a sandwich. One surface of each sandwich associates to form a barrel structure when a light and heavy chain comes together. Supported on this barrel scaffold the six CDRs pack together in the tertiary structure to form a relatively flat platform with a surface area of about 700 $Å^2$. All six CDRs are involved in the interaction in the anti-lysozyme complexes.

A1.1.3. Anti-lysozyme complexes

As the three dimensional structure of hen egg lysozyme has been determined to high resolution (Blake *et al.*, 1965), it serves as an ideal model antigen. Additionally structures of lysozyme from other species are available like human (Artimiuk and Blake, 1981) and turkey (Bott and Sarma, 1976). The antigenic response to HEL has been the subject of various studies. Smith-Gill *et al* (1984) and Harper *et al* (1987) have shown that the entire surface of HEL is potentially antigenic. Structures of three anti-lysozyme antibodies D1.3 (Amit *et al.*, 1986; Bentley *et al.*, 1990; Bhat *et al.*, 1990), HyHEL-5 (Sheriff *et al.*, 1987), HyHEL-10 (Padlan *et al.*, 1989) have been solved by X-ray crystallography. In the three anti- lysozyme antibodies the epitopes have been found to almost non-overlapping though a small overlap is seen between D1.3 and HyHEL-10. A striking feature of the interaction of these three antibodies with lysozyme is the shape complementarity of the two surfaces as revealed by the total exclusion of water

from the interface over an area of about 700 $Å^2$ for both the antigen and antibody. Water exclusion plays an important role in stabilising the complex. The structures of a number of anti-peptide antibodies that cross react with native HEL like Gloop2 have also been determined (Jeffrey, 1989). The Gloop2 system has been studied in great detail in the laboratory.

A1.1.4 The Gloop2 system

Gloop1 to 5 are a group of monoclonal antibodies specific for the loop region of HEL, which comprises of residues 57-84. The antibodies were raised against a peptide-bovine serum albumin conjugate and crossreact with native HEL (Darsley and Rees, 1985a and b). Gloop2 binds the loop antigen with an affinity of 10^8 M^{-1} and to HEL with an affinity of 10^7 M^{-1} (determined by solid phase RIA) [Roberts *et al.*, 1987].

The epitopes recognised by Gloop1 and Gloop2 are indistinguishable (Darsley and Rees, 1985a). The epitope on HEL has been determined by serological mapping of Gloop2 using a panel of avian lysozymes and it has been further refined by NMR spectroscopy. A catalytic site was designed in the combining site of Gloop2 using the aspartic protease active site as a basis for the model. The site consists of 2 aspartate residues at positions 32 (L1) and 92 (L3) in the light chain. Two single mutants Tyr32Asp, Leu92Asp and the double mutant Tyr32AspLeu92Asp were made by site directed mutagenesis (Hilyard, 1991).

A1.1.5 Fv expression

For X-ray crystallographic studies, the double Asp (Tyr32AspLeu92Asp) mutant was expressed as Fv fragments in E.coli and purified by affinity chromatography. The Fv fragments are non-covalently associated heterodimers of the H and L chain variable domains.

Their stability can be improved if they are produced as a single chain by inserting a linker peptide between the H and L chains (ScFv). ScFv have the specific antigen binding affinity comparable to that of unmodified Fab fragments and have been used in tumour imaging and in the form of recombinant immunotoxins for tumor therapy. Fv fragments can also be stabilised by introducing a disulphide bond between the two chains (Glockshuber *et al.*, 1992).

Bacterial systems offer an alternative to mammalian expression when glycosylation is not required. The E.coli secretion expression system was chosen mainly because of the speed and convenience of E.coli molecular biology manipulations. It also allows characterisation of many variants in a relatively short time. This expression system targets the recombinant protein to the periplasm of the bacterial cell, where the oxidising environment allows for the formation of disulphide bonds. In the case of antibody fragments, the probability of producing correctly folded protein is increased. The level of proteases in the periplasm is lower than in the cytoplasm reducing the risk of degradation of the recombinant protein.

The mutant Fv was purified by affinity chromatography. This method selects for antigen binding function and enables a large volume of crude supernatant to be concentrated into a small volume of pure protein. Thus it selects for correctly

folded functional Fv with sufficient purity for crystallographic studies.

A1.2 Materials and Methods

A1.2.1 Optimised expression and secretion conditions

A1.2.1.1 Induction of expression

A freshly transformed bacterial colony containing the pelB-GL2 Fv plasmid was used to inoculate an overnight culture. The culture was diluted to 1% in L-broth containing ampicillin (100 µg/ml) and 1% glucose and grown at 250 rpm, 30°C for 15 hours. The cells were pelleted by centrifugation using a GSA rotor in a Sorval RC-5 centrifuge at 7000 rpm for 10 minutes and the supernatant discarded. The cells were then gently washed in half the culture volume of 50mM NaCl. The 50mM NaCl wash was repeated and the cells resuspended in L-broth containing ampicillin (100 µg/ml) and 1mM IPTG (Boehringer). The culture was incubated for a further 17 hours at 250 rpm, 30°C.

After the induction period, the culture was centrifuged in the GSA rotor at 10,000 rpm, 4°C for 30 minutes and the supernatant was harvested. Protease inhibitors were added to a final concentration of 1 µM leupeptin (Boehringer) and 0.1mM EDTA. The supernatant was filtered through 0.45 µM Durapore hydrophilic membrane (Millipore) to remove cell debris and stored at -20°C.

A1.2.1.2 Preparation of cell fractions

At time points during the induction, an aliquot of culture equivalent to 2 OD_{550nm} units of cells was taken and centrifuged at 13000 rpm for 15 minutes at 4°C. The supernatant was removed and stored at -20°C (the supernatant fraction). The cell pellet was resuspended in 500 µl of 20% w/v sucrose, 30 mM Tris-HCl pH 8.0, and 50 µl stored at -20°C (whole cell fraction). A 2 µl aliquot of 0.25M EDTA pH 8.0 was added to the remaining 450 µl and incubated at 30°C for 15 minutes. The solution was transferred to a cold microfuge tube and centrifuged at 13000 rpm for 5 minutes at 4°C. The supernatant was discarded and the pellet resuspended in 500 µl ice cold water by brief vortexing. After incubation at 4°C for 10 minutes and centrifugation for 5 minutes, the supernatant was removed and stored at -20°C (the periplasmic fraction).

The Fv was purified from the crude supernatant by affinity chromatography.

A1.2.2 Purification of DD-Fv on a lysozyme affinity column

A 100ml sample of crude supernatant from a W3110/pelB-GL2DD Fv culture was loaded onto a prewashed 3ml lysozyme Sepharose column and washed with 10 volumes of phosphate buffered saline (PBS). The protein was eluted with 50mM diethylamine and samples separated by SDS-PAGE. Western blotting with anti-VL antisera showed that there was immunoreactive material in the breakthrough and the wash fractions in addition to the diethylamine eluted fractions.

The affinity of DDFv for lysozyme was not high enough for affinity chromatography under these conditions. Therefore purification of DDFv by loop peptide affinity chromatography was tested.

A1.2.3 Preparation of loop peptide

The loop peptide, a loop fragment of HEL consisting of residues 57-84 was synthesised by a modification of the method used by Hilyard, 1991.

5g of hen egg white lysozyme (Sigma) was dissolved in 330 ml 5% formic acid. It was then subjected to proteolytic digestion with pepsin at 37°C. The reaction was stopped by the addition of an equal volume of water after 16 hours and then freeze dried. The freeze dried material was resuspended in 15 ml 1% formic acid and spun at 6000 rpm for 10 minutes to get rid of high molecular weight contaminants.

The supernatant was then loaded onto a 100 ml G-25 medium Sephadex column pre-washed with 1% formic acid. The peptide is eluted with 1% formic acid and its purity was checked on a fast gel and by high pressure liquid chromatography (Figure A1.1).

A1.2.4 The use of blocked Sepharose to remove non-specific contaminants

Sepharose matrix adsorbs proteins non-specifically, often due to charged groups from the glycine blocking agent (Pharmacia LKB instructions). A blocked

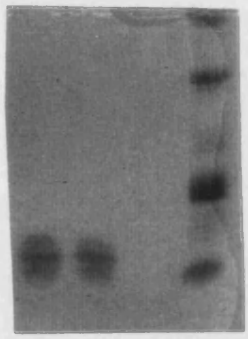


Figure A1.1 Pure loop peptide is seen in lanes 1 and 2 along with the protein marker in lane 4.

Sepharose column was used to remove non-specific proteins. The Sepharose was blocked in 0.2M glycine pH 8.0 overnight at 4°C and washed with acetate buffer (pH 4).

A 800ml sample of crude supernatant from a W3110/pelB-GL2Fv DD culture, adjusted to 0.25M NaCl, was loaded onto a prewashed 5ml loop-Sepharose column. The column was washed in 10-20 volumes of low salt wash (Tris HCl pH 8.0 containing 0.25M NaCl) and 3 volumes of high salt wash (Tris HCl, pH 8.5 containing 0.5M NaCl). The protein was eluted in 50mM diethylamine. Under these conditions it was possible to purify the mutant Fv to over 95% purity. The final yield of Fv purified protein ranged from 1-2 mg/litre estimated by absorbance at 280nm [Gloop 2Fv molar extinction coefficient is 38300 $M^{-1}cm^{-1}$ (Field, 1988)].

A1.2.5 Analysis of the VL polypeptide by Western blotting

Protein samples were separated by SDS-PAGE (as described in Appendix 3) and transferred onto a nitro cellulose membrane (Schleicher and Schuell) in contact buffer (40mM Glycine, 50mM Tris, 20% ethanol) using graphite electrodes (Multiphore apparatus, LKB). The membrane was blocked in low salt buffer (LSB: 10mM Tris, pH 7.6; 0.9 % w/v NaCl; 0.05 % w/v Tween 20) for 1 hour at room temperature and then incubated with the first antisera (rabbit anti-GL2VL IgG prepared by H. Field) at 1/200 dilution in LSB overnight at 4oC. After washing twice for 10 mins in LSB, the membrane was incubated with 1/2000 dilution of goat anti-rabbit IgG alkaline phosphatase conjugate (Sigma) in LSB for 1 hour at room temperature. The membrane was washed as described above and then

developed in 75mM Tris-HCl, pH 9.6; 4mM MgCl2; 0.005% w/v nitroblue tetrazolium; 0.0025% w/v 5-bromo-4chloro-3-indolyl phosphate.

A1.3 Results

A1.3.1 Cellular distribution of the VH and VL polypeptides

Western blotting of bacterial extracts with either anti-GL2VL or anti-GL2VH antisera verified the expression and secretion of both the VH and VL polypeptides in BMH 71.18 cells containing the pel B GL2 DDFv vector.

Cell fractions were taken from a 500ml culture of BMH 71.18/ pel B-GL2DDFv. The culture was induced with 1mM IPTG at 0.5 OD_{550nm} units and samples taken at 0 and 17 hours after induction. Cell fractions were prepared as described in Section 1.2. The samples were separated by SDS-PAGE (Figure A1.2) and visualised by Western blot with anti-VL antisera (Figure A1.3)

VH and VL polypeptides were seen by Coomassie staining in the cell and periplasmic fractions 5 hours after induction. There was no detectable immunoreactive material in the periplasmic fraction after induction; this may be due to the periplasmic contents passively leaking into the supernatant through holes in the cell wall. VH and VL were detected in the supernatant by Western blot analysis after induction for 17 hours.

A1.3.2 The effect of host strain on the expression level of VH and VL polypeptides.

Parallel inductions were performed with the strains W3110 and BMH 71.18 to compare the relative levels of VH and VL expression. The yield of secreted product from the two strains was similar.

A1.3.3 Purification of the loop peptide

The Fv was purified from the crude supernatant by affinity chromatography on a loop peptide column. This loop peptide was prepared by a modification of a method used by Hilyard K.L. (DPhil thesis, University of Oxford). This method was chosen as it enables the protein to be purified with antigen binding function and a large volume of crude supernatant can be concentrated into a few millilitres of pure protein (Figure A1.4).

A1.4 Discussion

The yield of purified mutant Gloop2 DDFv ranged between 1-2 mg/litre of culture. This variation may be due to losses during purification as yields were only estimated after affinity chromatography by Coomassie staining and absorption at OD_{280nm} Inefficient induction of expression or plasmid loss during the induction could also account for this variation.

For Fv production, levels in the 100-500 mg/l range are achievable in E.coli compared to low mg/l levels from mammalian expression systems. The double Asp

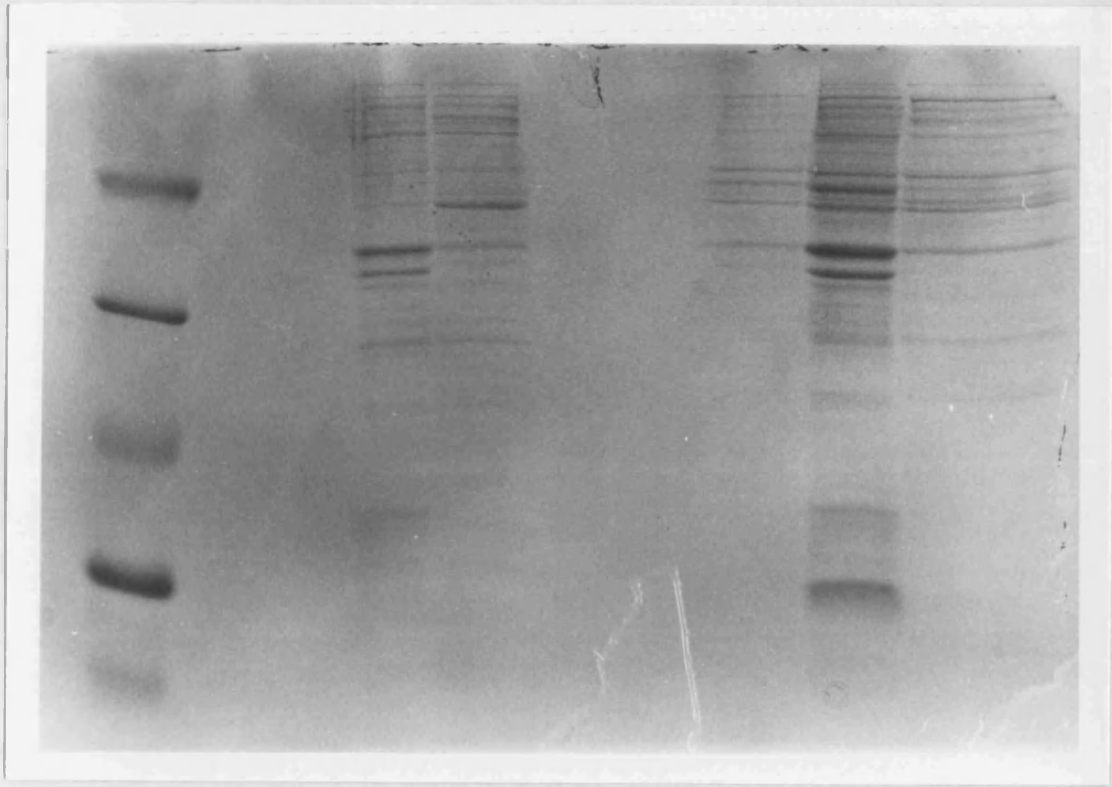


Figure A1.2 Coomassie staining showing supernatant fraction (0 and 17 hours in lanes 3 and 8 respectively), washed cell fraction (0 and 17 hours in lanes 4 and 9 respectively), periplasmic fraction (0 and 17 hours in lanes 7 and 10 respectively). Only the supernatant fraction after induction for 17 hours had Fv protein.

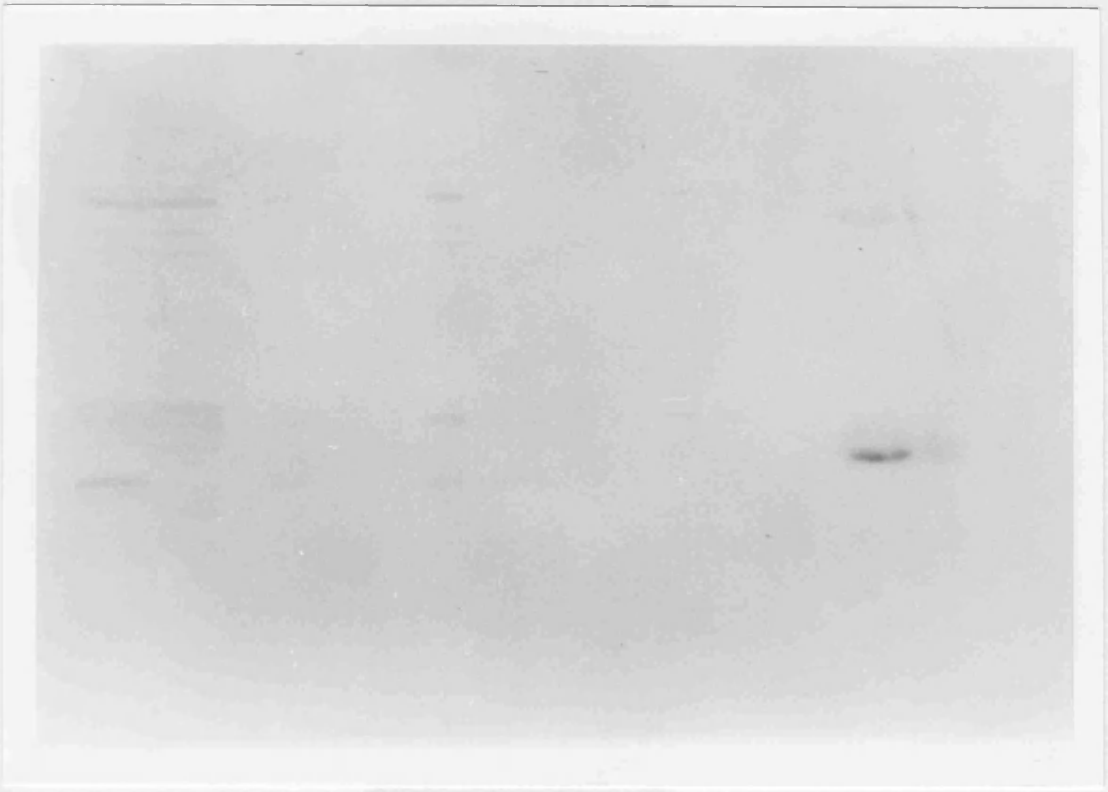


Figure A1.3 : Western blot showing supernatant fraction (0 and 17 hours in lanes 3 and 8 respectively), washed cell fraction (0 and 17 hours in lanes 4 and 9 respectively), periplasmic fraction (0 and 17 hours in lanes 7 and 10 respectively) with anti-VL antisera. Only the supernatant fraction after induction for 17 hours had Fv protein.

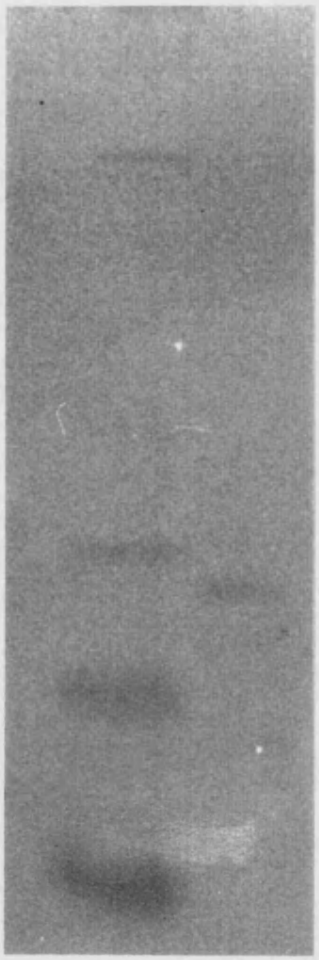


Figure A1.4 Coomassie staining of pure Fv protein in lane 2 along with protein marker in lane 1.

(Tyr32AspLeu92Asp) mutant antibody was shown to retain binding affinity for peptide. Catalytic assays with both iodinated and unlabelled peptide antigen indicated that modification of peptide occurs on incubation with the mutant Fv. A different modification to the unlabelled peptide was observed in the wild type Fv (Hilyard, 1991). X- ray crystallographic studies could be used to investigate the nature of peptide modification. Large quantities of mutant Fv had to be produced for crystallographic studies. Large scale production of Gloop2 DDFv will be attempted in a 5 to 7 litre bench top fermenter. Initial experiments with similar Fv proteins have yielded approximately 10 mg/ml by fermentation. The pure antibody protein produced can be used for X-ray crystallographic and NMR studies.

APPENDIX-2

KM vs Priess TCR Vα and Vβ gene cloning

A2.1 V_β gene sequence analysis

One of the original aims of this project was to analyse TCR V_α and V_β gene usage in peptide specific DR4 Dw4 restricted T cell clones. A strategy to sequence TCR V_α and V_β genes from the T cell line KM vs Priess was devised. The materials and methods used in the study of KM vs Priess cell line are outlined in this section. However these methods could not be extended to the study of peptide specific DR4 Dw4 restricted T cell clones as they were contaminated with mycoplasma.

KM vs Priess cell line and a DR4 Dw4 restricted T cell clone HF53 was kindly given by David Sansom from Bath Institute of Rheumatic Diseases.

A2.1.1 RNA extraction

Efficient extraction of total RNA is a pre requisite for subsequent cDNA synthesis and cloning. A major technical problem in the isolation of RNA is the

inhibition of RNAse activity. To minimise the deleterious effects of RNAses, all solutions and equipment for RNA work were treated with diethyl pyrocarbonate (DEPC). Electrophoresis tanks were cleaned with detergent solution, rinsed in water, dried with ethanol and then filled with a solution of 3% hydrogen peroxide for 10 minutes at room temperature. The tanks were then rinsed with DEPC treated water.

Many approaches to RNA isolation involve the simultaneous disruption of tissue and denaturation of all cellular proteins including RNAses. Rate of protein denaturation must exceed the rate of RNA hydrolysis. Most rapid denaturation of proteins involves the use of guanidine thiocyanate and 2 mercapto ethanol during tissue disruption. Guanidine thiocyanate in association with N-lauryl sarcosine acts to disrupt nucleoprotein complexes allowing RNA to be released into solution and isolated free of protein. Intact RNA is purified from contaminants by phenol/chloroform extraction based on the rapid one-step procedure of Chomczynski and Sacchi (1987). RNA selectively partitions into the aqueous phase free from DNA and protein and is easily concentrated by precipitation with iso-propanol. LiCl precipitations can result in the loss of RNA smaller than 5.8S and carryover of Li^+ salts can inhibit subsequent cDNA synthesis reactions. Various methods used for the extraction of RNA are outlined below.

A2.1.1.1 RNA extraction with guanidinium thiocyanate

Lysed T cells were resuspended in 5 mls of Solution D and 20 mls of pre chilled Solution I (guanidinium thiocyanate in buffer containing tris-HCl and

EDTA, pH 7.4; Solutions I, II,III and IV were supplied with Amersham's RNA extraction kit) was added in a tissue disruption tube. The cells were sonicated in 40 second bursts at 10 watts until the solution was less viscous. To the homogenate 0.3 volumes ethanol was added and the mixture was incubated at -20°C for 10 minutes. After centrifugation at 0°C for 5 minutes at 10000 rpm, the pellet was resuspended in 10 ml pre chilled Solution I. The cells were sonicated again and after the addition of 30 mls of Solution II (LiCl in water) the mixture was incubated at 4°C overnight. After centrifugation at 10000 rpm at 4°C for 90 minutes the supernatant was discarded. The pellet was resuspended in 35 ml Solution III (LiCl and urea in water) by alternate vortexing for 30 seconds and incubation at 40°C. After centrifugation at 10000 rpm for 60 minutes at 4°C the protein film was discarded and 5 mls of Solution IV (RNA buffer) was added to the pellet. After incubation at -20°C for 2 hours the pellet was resuspended by alternate vortexing for 30 seconds and incubation in 40°C water bath. The solution was then extracted once with phenol and twice with $CHCl_3$. RNA was precipitated with 2 volumes ethanol and the pellet was resuspended in 1 ml DEPC water. A quicker method was necessary for the routine extraction of RNA.

A2.1.1.2 RNA extraction using phenol chloroform isoamyl alcohol extraction

10^6 T cells from the DR4 Dw4 restricted T cell clone were resuspended in 500 µl of 4M guanidinium thiocyanate (Sigma G-6639); 25mM sodium citrate,

pH 7; 0.5% sarcosyl; 0.1 M 2 mercapto ethanol (solution D). 50 µl 2M sodium acetate, pH 4; 500 µl phenol (water saturated); 100 µl $CHCl_3$-isoamyl alcohol (49:1) were sequentially added and the mixture was shaken vigorously for 10 seconds. After cooling the mixture on ice for 15 minutes it was spun in a microfuge for 10 minutes. The aqueous layer was then extracted twice with an equal volume of $CHCl_3$-isoamyl alcohol. RNA was precipitated by the addition of an equal volume of isopropanol and an overnight incubation at -20°C. The pellet was resuspended in 0.3 volumes of solution D and after another isopropanol precipitation the pellet was washed twice with 70% ethanol and once with absolute ethanol. The pellet was then resuspended in 50 µl DEPC water.

A2.1.1.3 RNA extraction by RNAzol method

10^6 cells from the DR4 Dw4 restricted T cell clone were resuspended in 1.1 mls Phosphate Buffered Saline (PBS). The cells were then lysed by the addition of 0.2 ml of RNAzol (Biotecx Lab, Inc). RNA was solubilised by passing the lysate several times through the pipette. The homogenate was then extracted with $CHCl_3$. To the aqueous layer, an equal volume of isopropanol was added and the mixture was incubated at 4°C for 15 minutes. RNA was pelleted by centrifugation at 13000 rpm for 15 minutes at 4°C. The RNA pellet was washed once with 75% ethanol by vortexing and subsequent centrifugation for 8 minutes at 7500 rpm at 4°C. The pellet was dried under vacuum for 15 minutes and resuspended in 1mM EDTA pH 7. Isolated RNA exhibited a A_{260}/A_{280} ratio of 1.7-

2.0. A lower ratio is indicative of contamination with guanidine thiocyanate carried over during the subsequent steps.

This method was used for all subsequent RNA extractions.

A2.1.1.4 Gel analysis of RNA transcripts

The quality of RNA transcripts was assessed by loading a 5 µl aliquot of the sample on a 1% HCHO gel along with ribosomal RNA markers (Figure A2.1) RNA transcripts (0.5-2 µg) were resuspended in 1 X MOPS [3-(N-morpholino) propane sulphonic acid] buffer (20 mM MOPS; 5 mM sodium acetate, pH 5.0; 1mM EDTA, pH 7.0); 2.2 M formaldehyde and 70% (v/v) deionised formamide. The samples were denatured by heating at 65°C for 15 minutes followed by rapid cooling on ice. After the addition of 1 µl of 1 mg/ml ethidium bromide and 0.1 volume of gel loading dye (0.025% xylene cyanol; 0.025% bromo phenol blue; 4% glycerol) the samples were loaded immediately onto denaturing gels containing 1.0% agarose and 3.7% formaldehyde in 1 X MOPS buffer. The samples were separated by electrophoresis at a limiting current of 75 mA for 3-4 hours.

RNA was visualised by UV fluorescence as in the case of DNA. The 28S and 18S eukaryotic ribosomal RNAs exhibit a near 2:1 ratio of ethidium bromide staining in undegraded samples. In degraded samples this ratio will be reversed as the 28S ribosomal RNA is characteristically degraded to an 18S-like species.

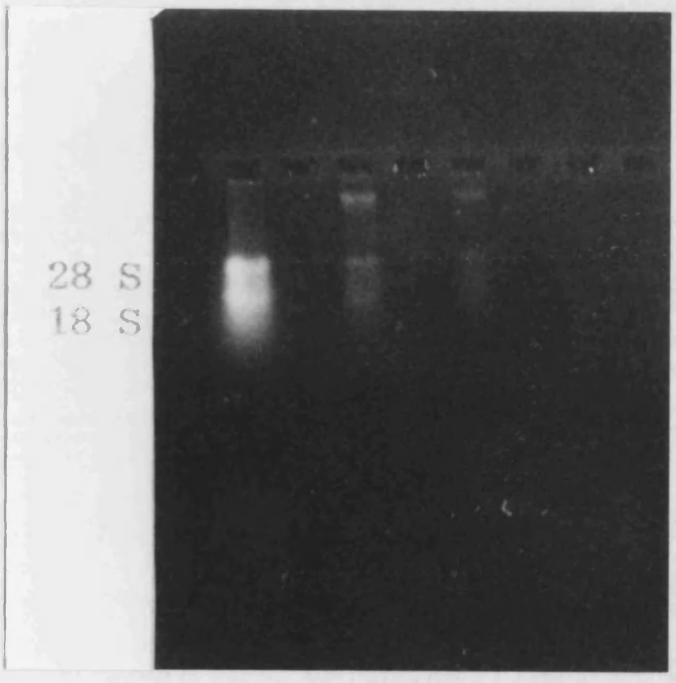


Figure A2.1 Total RNA from KM vs Priess cells seen in lanes 3 and 5 along with the marker in lane 1

A2.1.2 cDNA synthesis

The first strand of cDNA can be synthesised using an enzyme RNA-dependent DNA polymerase, the reverse transcriptase (RT). This enzyme can synthesise a complementary DNA strand initiating from a primer using either single-stranded RNA or DNA templates. Two different forms of reverse transcriptases are available commercially, Avian RT and Murine RT. Moloney Murine Leukemia Virus Reverse Transcriptase (M-MuLV RT) has much lower RNase H activity than Avian Myeloblastosis Virus Reverse Transcriptase (AMV RT), significant especially when synthesising cDNA from long mRNAs. Intact mRNA template remaining after cDNA synthesis can interfere with RT-PCR by competing with the cDNA as PCR template. Both reverse transcriptases lack 3' H 5' exonuclease function. First strand cDNA synthesis may be accomplished by extension with random hexamers, the downstream primer or thymidylic acid oligomers like oligo $(dT)_{18}$ and oligo $(dT)_{12\text{-}18}$.

Random hexamers are not preferred for reverse transcription as multiple primers bind to a single RNA creating either a cDNA that is prematurely terminated or one that is fully extended but consisting of several unlinked fragments that will become dissociated during PCR. For cloning of cDNAs, the most frequently used primer is Oligo $(dT)_{12\text{-}18}$ which binds to the poly(A) tract at the 3' terminus of eukaryotic cellular mRNA molecules.

A2.1.2.1 Comparison of reverse transcription reaction conditions

The relative efficiency of different available reverse transcriptases and

buffer systems was compared in order to choose the optimal reaction conditions for RACE. The different protocols were tested by RT-PCR using total RNA extracted from a T cell line KM vs Priess as a control.

Total RNA was washed with 75% EtOH, resuspended in DEPC treated water at 0.2 µg/µl. 5 µl aliquots of RNA were taken in microfuge tubes in the presence of the appropriate 1 × reverse transcription buffer and primers. All reactions were terminated by the addition of 4 µl of 0.5 M EDTA, pH 8.

8 µl aliquot of each reverse transcription reaction was added to a 50 µl PCR mix containing the V_β gene specific primer set AN Poly C and C_β. PCR amplification was carried out as described in Section 2.4.4. A 10 µl aliquot of each PCR reaction was analysed on a 1% agarose gel. The intensity of the 600 bp PCR product generated from the V_β gene coding sequence was used as an indication of efficiency of each reaction. This comparison was not designed to be quantitative or comprehensive but rather to establish the most appropriate reaction conditions using available reagents.

C_β gene specific primed cDNA synthesis

Total RNA extracted from a T cell line KM vs Priess was used for cDNA synthesis in the presence of 0.5 µg of a C_β gene specific primer. A C_β gene specific primer

5' GGGACACTTAAGGGTAGAGCAGGTACT 3' (Tm = 82°C)

was designed and synthesised on an Applied Biosystems DNA Synthesiser (described in detail in Section 2.4.1). The reaction was performed in the presence of 0.5 mM dNTPs; 10 mM DTT; 75 mM KCl; 50 mM Tris HCl, pH 8.3; 3 mM MgCl2 and 200 units of M-MLV RT in a reaction volume of 20 µl. This reaction was

carried out at 37°C for 60 minutes.

Oligo dT primed synthesis of first strand of cDNA

Protocol 1

Reverse transcription was performed using Superscript with Gibco BRL supplied buffer for 60 minutes at 37°C. This reaction was carried out in the presence of 50 mM Tris HCl, pH 8.4; 50 mM KCl; 2.5 mM MgCl2; 100 µg/µl BSA; 0.5 µg Oligo $(dT)_{12\text{-}18}$ and 200 units of Superscript enzyme. The reaction volume was 30 µl.

Protocol 2

cDNA synthesis was carried out in the presence of 50 mM Tris HCl, pH 8.3; 50 mM KCl; 10 mM DTT; 0.5 µg Oligo $(dT)_{12\text{-}18}$ and 200 units of M-MLV RT (Gibco BRL) for 60 minutes at 37°C. The reaction volume was 30 µl.

Protocol 3

1 µg of total RNA from the T cell line KM vs Priess was used for Oligo dT primed cDNA synthesis using 0.5 µg of Oligo $(dT)_{12\text{-}18}$ annealed at 70°C for 10 minutes. Reactions were performed using 0.5 mM dNTPs; 10 mM DTT; 75 mM KCl; 50 mM Tris HCl (pH 8.3); 3 mM $MgCl_2$ and 200 units of Moloney Murine Leukemia Virus Reverse Transcriptase (BRL) in a final volume of 20 µl, at 37°C for 60 minutes. The reaction was terminated by the addition of 4 µl of 0.5 M EDTA, pH 8.0. The progress of this reaction was monitored by incorporating 10 µCi α-^{32}P dCTP radiolabel. This method was used for all subsequent cDNA synthesis.

A2.1.2.2 cDNA quantitation

DE-81 filters are positively charged and will strongly adsorb and retain nucleic acids including oligonucleotides that are too small to be precipitated efficiently with tri chloro acetic acid (TCA). Unincorporated nucleotides stick less tightly to the filters and are removed by washing the filters extensively in 0.5M Na_2HPO_4, pH 7. The efficiency of this reaction can be calculated by measuring the proportion of radioactive precursor incorporated into the cDNA. This can be done by two methods namely 1) Differential precipitation of the nucleic acid products with TCA and 2) Differential adsorption of the products onto positively charged surfaces eg DE-81.

1 µl of cDNA was spotted onto the centre of two DE-81 filters. One filter was used to measure the total amount of radioactivity in the reaction. This filter was washed thrice with 0.5M Na_2HPO_4, pH 7.0 for 2 minutes. It was then rinsed with water for a minute followed by a wash in 70% ethanol for 4 minutes and then dried. Each of the filters (washed and unwashed) was inserted into scintillation vials and the amount of radioactivity on each filter was measured. ^{32}P can be determined on dry filters by Cerenkov counting (in the ^{3}H channel of a liquid scintillation counter). The amount of radioactivity on the unwashed filter can be compared with the amount of radioactivity on the washed filter and the % incorporation calculated (Table A2.1).

A2.1.2.3 cDNA purification

22 µl cDNA was loaded onto a 3 ml G-50 Sephadex column washed with

Fraction	cpm
1	65.7
2	16
3	27
4	3100.7
5	32098
6	7151.7
7	9618.7
8	190432.7
9	1981295.6
10	1897432.5
11	1467654.2
12	1298603.5
13	299814.7
14	70415.0
15	55663.7
16	29056.0
17	28795.0
18	23836.3
19	24188.3
20	15027.7

Table A2.1 Counts per minute in the fractions collected from a Sephadex G-50 column during KM vs Priess cDNA purification as described in section A2.1.2.2. Fraction 5 contained the cDNA.

TEN buffer (10 mM Tris HCl; 1 mM EDTA, pH 8; 100 mM NaCl). 20 fractions (100 µl each) were collected and the radioactivity was measured in a Cerenkov counter (Table A2.1). Fraction 5 contained the cDNA.

A 4 µl aliquot of the sample was run on a 1% alkaline gel containing 2.5 ml 1M NaOH and 100 µl 0.5 M EDTA, pH 8.0. The gel was fixed in 7% TCA overnight. It was then dried and subjected to autoradiography at -70°C (Figure A2.2).

A2.1.3 G-Tailing with Terminal deoxynucleotidyl Transferase

Terminal deoxynucleotidyl transferase (TdT) catalyses the repititive addition of mononucleotides from a dNTP to the terminal 3'-OH of a DNA initiator accompanied by the release of inorganic phosphate. This enzyme was first purified from calf thymus by Bollum (1958).

It provides a unique method for the labelling of the 3' termini of DNA with ^{32}P for subsequent utilisation in hybridisation assays. The enzyme to substrate ratio is critical for obtaining uniform addition of labelled nucleotides. Lower ratios produced probes of varying lengths and reduced incorporation rates. All types of double stranded DNA ends (blunt ends, 3'-overhangs, 5'-overhangs) can be tailed with TdT in the presence of a cobalt-based buffer. The reaction volume was made upto 100 µl with water and the cDNA was purified by 3 precipitations with 1 volume 4 M ammonium acetate and 3 volumes of ethanol. The pellet was washed with 70% ethanol and spun for 15 minutes. It was then dG tailed using either of the two methods.

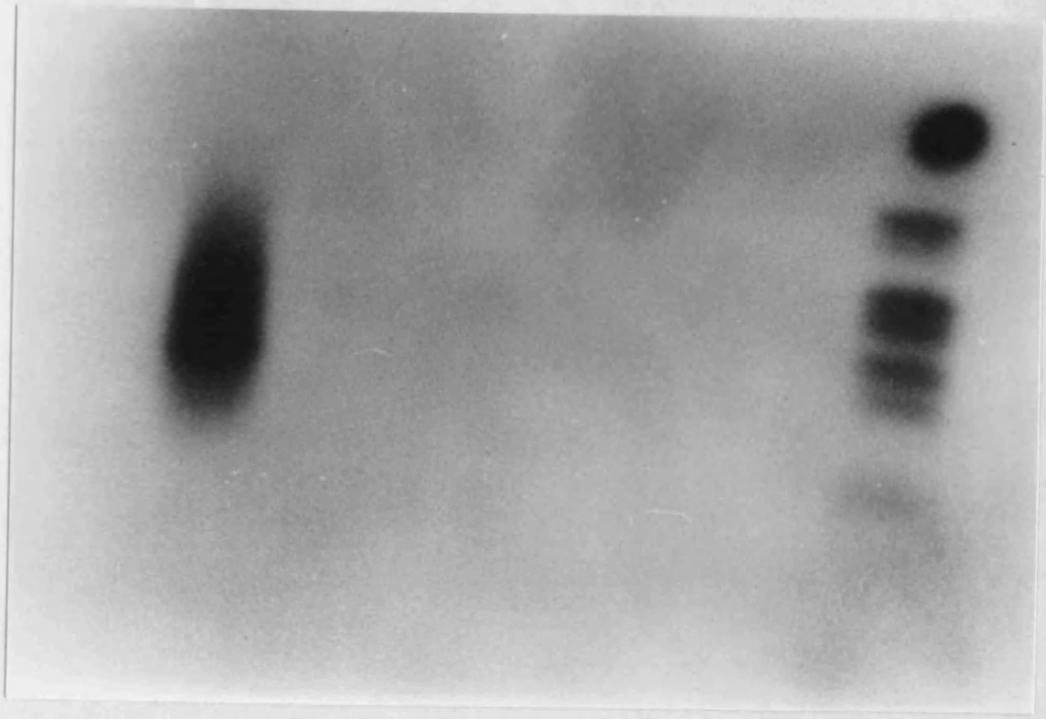


Figure A2.2 Autoradiograph of KM vs Priess cDNA seen along with ϕX 174 RF DNA marker.

Protocol 1

The cDNA pellet was resuspended in 37.775 µl water and dG tailed in the presence of 0.5 M potassium cacodylate (pH 7.2); 10 mM $CoCl_2$; 1 mM DTT; 24.5 µM dGTP and 15 Units TdT (Gibco BRL) in a final volume of 50 µl for 30 minutes at 37°C. The tailed cDNA was then diluted to a volume of 100 µl.

Protocol 2

The cDNA pellet was resuspended in 9.5 µl water and dG tailed in the presence of buffer with 1 M potassium cacodylate; 125 mM Tris HCl; 1.25 mg/ml bovine serum albumin, pH 6.6; 0.75 mM $CoCl_2$; 0.75 mM dGTP and 50 units TdT (Boehringer Mannheim) in a final volume of 20 µl for 15 minutes at 37°C. The tailed cDNA was then diluted to a volume of 100 µl.

Reverse transcription reactions were diluted prior to RACE to limit the amount of initial template for RACE to favour exponential amplification of specific sequence over non specific background. Excess nucleotides and RT primer have to be removed from the first strand reaction prior to RACE and therefore a stringent purification procedure was required. Contaminating proteins were removed by performing extractions with phenol and chloroform. Tris-buffered phenol, chloroform and isoamyl alcohol (25:24:1; v/v) were added to an equal volume of DNA. Phases were mixed by vortexing and separated by centrifugation at 13000 rpm for 5 minutes. The upper aqueous phase containing DNA was transferred to a new microfuge tube and then extracted twice with chloroform. The aqueous phase was used as template for Anchored PCR.

A2.1.4 RACE method

The RACE method (Rapid Amplification of cDNA ends) first described by Frohman (1988) has also been described as one-sided PCR or Anchored PCR. It was designed to rapidly produce, isolate and characterise unknown cDNAs. In theory, with the use of nested primers, transcripts can be amplified to generate a pure product containing the desired cDNA. It is quicker than other methods in that the reverse transcribed products can be analysed within 1-2 days after the experiment. However there is an enormous potentialfor the generation of artifacts by this method.

The most obvious artifact is the non specific amplification of cDNAs to generate gene products which have no sequence homology to the gene of interest. There are other PCR products which are undesirable either because they are truncated or because they are derived from contaminating genomic sequences. Truncated cDNA 3' ends are produced when the oligo $(dT)_{12-18}$ primer anneals to a A rich region in the transcript upstream of the poly (A) tail. This can be avoided to some extent by decreasing the concentration of the oligo $(dT)_{12-18}$ primer to 15 ng/reverse transcription.

For the 5' ends, a primer which binds inappropriately to the message of interest within the region to be amplified causes premature termination of the reverse transcription reaction at that site creating truncated ends. To overcome this problem, primers having a high A+T content can be used and the reverse transcription reaction can be carried out at 52°C. Primers with a high G+C content are more likely to produce artifacts.

Anchored PCR (An PCR) is a useful method for analysing RNA transcripts where the sequence at the 5' end is unknown. The procedure involves reverse transcription of mRNA using oligo $(dT)_{18}$ or a gene specific primer, homopolymer tailing of the cDNA with dGTP and then amplification by PCR using a non specific anchor oligo (dC) at the 5' end and a gene specific primer at the 3' end (Loh, 1989 and Frohman, 1988) [Figure A2.3].

If the oligo $(dT)_{18}$ primer is not completely removedfrom the cDNA prior to the homopolymer tailing step the reaction will fail even if the same cDNA can be amplified using 3' and 5' gene specific primers. (Jitra Kriangkum, 1992). Presence of the oligonucleotide does not interfere with the homopolymer tailing itself as all the cDNA was found to be tailed. The reaction is inhibited by dG tailed oligo $(dT)_{18}$. When a series of An PCR reactions with increasing concentrations of $(dT)_{18}$ and dG tailed $(dT)_{18}$ was carried out with specific 5' and 3' primers, the An PCR product disappeared in the presence of 1 ng or more of dG tailed oligo $(dT)_{18}$. 400 ng of oligo $(dT)_{18}$ had no effect on this reaction. Oligo $(dT)_{18}$ can be completely removed by column chromatography over Biogel A5M (Bio Rad).

A2.1.4.1 Design, synthesis and purification of oligonucleotides

Two primers C_β 3' and AN Poly C (Table A2.2) were designed for use in Anchor PCR. A Not I site and a Bgl II site were introduced into the AN Poly C and C_β primers respectively. These primers were synthesised on an Applied Biosystems DNA synthesiser using phosphoramidate chemistry. The oligonucleotide sequences were checked against the DNA sequences of human TCR β chain

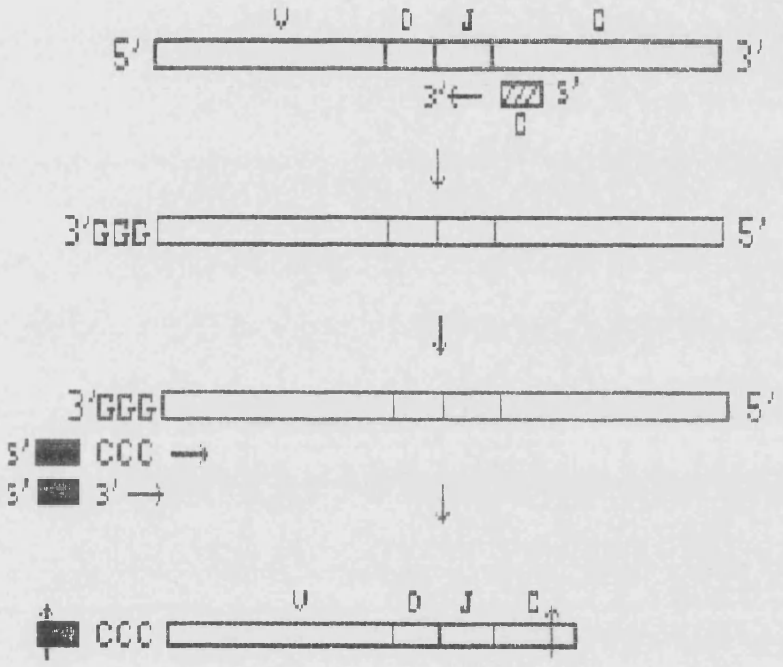


Figure A2.3 Schematic diagram of Anchored PCR.

Primer	Sequence	Tm (°C)
AN Poly C	5' CTATCTAGAGAGCTCGCGGCCGCCCCCCCCCCCC 3'	124
C_α 3'	5' CAGACAGATCTGTCACTGGAT 3'	62
C_β 3'	5' CGCGAATTCAGATCTCTGCTTCTGATG 3'	76
Amp 3'	5' GTGTGGGAGATCTCTGCTTCTGAT 3'	76
RADS	5' AGGACCTGAACAAGGTGTTCCCACCCGAGGTCG 3'	106
VBUN	5' GGGGAGCTAGCTAGCTAGCTAGTTATCCTTGGT 3'	102
CA-JBM	5' ATATCCAGAACCCTGACCCTGCCGTGT 3'	84
C_α 3'	5' GATAGATCTTAGAGTCTCTCAGC 3'	66

Table A2.2 Primers used in the study of TCR Vα and Vβ genes.

constant region exon 1 using an identity search program. Ammonia cleaves the oligo from the resin at room temperature. The protecting groups were removed from the terminus by incubating the resin in 0.88 N ammonia solution at 55°C for 16 hours. Ammonia was removed by drying the oligo under vacuum. Ethanol precipitated oligo can be recovered by a 10 minute spin at 10000 rpm followed by a rinse with 70% ethanol and then dried. Concentration of the oligo was determined spectrophotometrically.

A2.1.4.2 5'phosphorylation of oligo primers

C_β 3' and the AN Poly C primers were 5' phosphorylated. A 2 picomole sample of primer was mixed with 16 µl buffer (0.05 M Tris HCl, pH 7.6; 0.01 M MgCl2; 5 mM DTT; 0.1mM spermidine HCl; 0.1 mM EDTA, pH 8) and 7 units T4 polynucleotide kinase enzyme (Pharmacia) in a final volume of 20 µl. The reaction was performed at 37°C for 30 minutes and the kinase was inactivated by heating at 70°C for 10 minutes. The 5' phosphorylated primers were stored at -20°C.

A2.1.4.3 Control reactions for RACE

Control reactions were used to test the efficiency of reverse transcription and to ensure the validity of RACE products.

The efficiency of the cDNA reaction can be determined by using primers which amplify ubiquitously expressed mRNAs such as glyceraldehyde 3-phosphate dehydrogenase (G3PDH) or β actin.

A 8 µl aliquot of each reverse transcription reaction was taken in a 50 µl PCR mix containing 0.4 µM 5' and 3' cross species G3PDH specific primers, 0.2 mM dATP, 0.2 mM dCTP, 0.2 mM dGTP, 0.2 mM dTTP, 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$ and 2.5 Units Taq Polymerase (Perkin Elmer). The samples were subjected to PCR under the following conditions: 94°C, 45 seconds; 60°C, 45 seconds; 72°C, 2 minutes for 35 cycles followed by a time delay of 7 minutes at 72°C.

Cross species primers used to amplify a 0.45 kb G3PDH PCR product were:

5' primer

5' ACCACAGTCCATGCCATCAC 3' Tm = 62°C

and 3' primer

5' TCCACCACCCTGTTGCTGTA 3' Tm = 62°C

PCR products having a size of around 0.45 kb can be obtained using cross species G3PDH specific primers. A negative control containing no reverse transcriptase in the reaction mixture was also included along with each transcription reaction. This control reaction was used to determine whether a given RACE product is of genomic DNA or cDNA origin. Another control reaction omitting RNA from the reaction mixture was included to ensure that there was no contamination with PCR products or plasmid DNA.

A2.1.4.4 Anchored PCR

8 µl of the tailed cDNA was subjected to Anchor PCR using 2.5 units

of Taq Polymerase (Perkin-Elmer) and a Perkin Elmer Cetus Thermocycler under the following conditions: 94°C, 5 minutes; 55°C, 2 minutes; 72°C, 10 minutes for one cycle followed by 94°C, 45 seconds; 55°C, 2 minutes; 72°C, 10 minutes for 30 cycles. The reaction was performed using 200 µM dNTPs, 50 mM KCl, 10 mM Tris HCl, 0.1% Tween 20 and 1.5 mM $MgCl_2$. The primers used were Anchor Poly C and C_β 3' (Table A2.2) at a concentration of 8.9 pm. The volume was made up to 50 µl.

A 10 µl aliquot of the sample was run on a 1% agarose gel along with ϕX 174 RF DNA marker (Figure A2.4).

A2.1.5 PCR using a V_β specific universal primer

For PCR amplification of TCR β cDNA without prior knowledge of its V_β sequence, a nucleotide sequence of 5 bases (TGGTA) was chosen as a target for upstream PCR primer. This sequence coding for tryptophan and partly tyrosine, is highly conserved among all V_β genes known to date. A 18 base long V_β specific universal PCR primer designated VBUN containing the above 5 bases at the 3' end was designed. The other portion of VBUN contained 2 thymines, various degrees of base degeneracy designed on the basis of published V_β sequences, and a stretch of 4 guanines at the 5' end for stabilisation of annealing.

The oligonucleotide VBUN (Table A2.2) was synthesised on an Applied Biosystems DNA Synthesiser using phosphoramidate chemistry. The oligonucleotide sequence was checked against the DNA sequences of human TCR β chain CDR1 using an identity search program. Ammonia cleaves the oligo from the

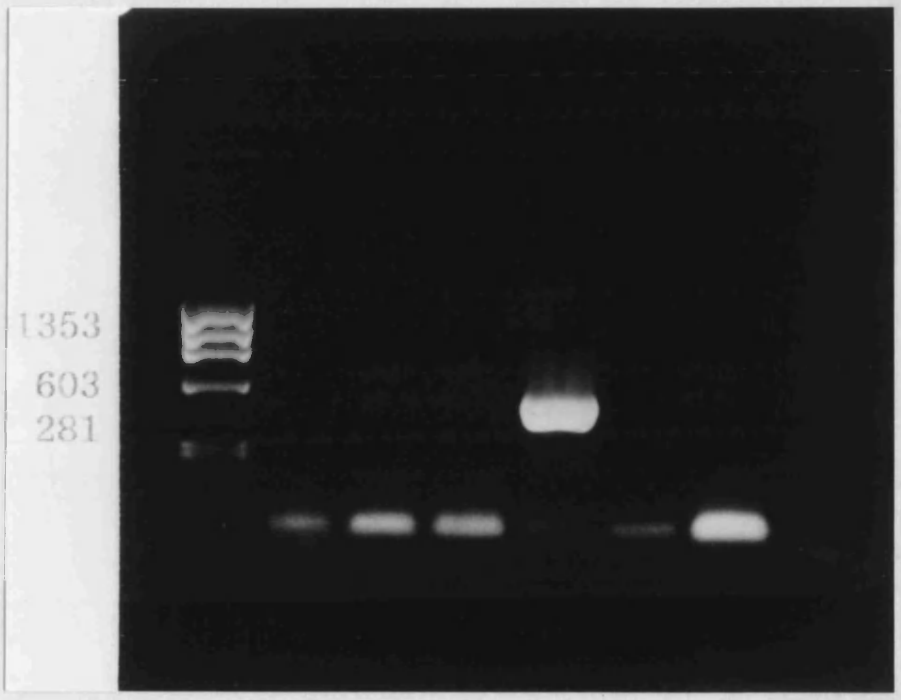


Figure A2.3 Anchored PCR product of TCR Vβ gene from KM vs Priess cells seen in lane 5 along with ϕX 174 RF DNA marker in lane 1.

resin at room temperature. The protecting groups were removed from the terminus by incubation at 55°C for 16 hours. Ammonia was removed by drying the oligo under vacuum. Ethanol precipitated oligo can be recovered by 10 minute spin at 10000 rpm followed by a rinse with 70% ethanol and then dried. Concentration of the oligo was determined spectrophotometrically.

4 µl of the tailed cDNA was subjected to PCR using 2.5 units of Taq Polymerase (Perkin-Elmer) and a Perkin Elmer Cetus Thermocycler under the following conditions: 94°C, 1 minute; 55°C, 1 minute; 72°C, 3 minutes for 25 cycles. The reaction was performed using 200 µM dNTPs, 50 mM KCl, 10 mM Tris HCl, 0.1% Tween 20 and 1.5 mM $MgCl_2$. The primers used were VBUN and C_β 3' (Table A2.2) at a concentration of 8.9 pm. The volume was made up to 50 µl.

A 5 µl aliquot of the sample was run on a 1% agarose gel along with ϕX 174 RF DNA marker.

A2.1.6 Purification of the PCR product

A number of methods were tried for the purification of the PCR product.

A2.1.6.1 Spun columns

1 µg of the PCR product resuspended in 100 µl water was loaded onto a 1 ml column of Bio Rad P-30 (Bio gel) pre equilibrated with 10 mM Tris HCl, pH 8; 0.5 mM EDTA, pH 8. A 100 µl fraction was collected after centrifugation at

1600 rpm for 4 minutes.

A 10 µl aliquot of this fraction sample was run on a 1% agarose gel along with ϕX 174 RF DNA marker.

.A2.1.6.2 Crush-Soak method

The amplification product was extracted with an equal volume of phenol: chloroform and subsequently with an equal volume of chloroform. After the addition of 0.5% SDS; 10 mM Tris HCl, pH 8; 5 mM EDTA and 50 µg Proteinase K the aqueous layer was incubated at 37°C for 30 minutes. The enzyme was then inactivated by heating the mixture at 68°C for 10 minutes. This mixture was extracted with an equal volume of phenol and subsequently with an equal volume of phenol:chloroform. The aqueous layer was then precipitated with 3mM spermine by incubation at room temperature for 10 minutes followed by centrifugation at 13000 rpm for 15 minutes. The pellet was resuspended in 300 µl 0.5 M NaCl and incubated at room temperature overnight. DNA was precipitated with ethanol and the pellet was rinsed with 70% ethanol. The pellet was then resuspended in 20 µl water.

The product was subjected to double digestion with Not I and Bgl II in the presence of 100 µg BSA. This product was then run on a 2% Nusieve GTG agarose gel in TAE buffer. DNA running at around 600 bp was eluted from the gel by crush-soak method. The PCR product was chopped with a Schwann Morton surgical blade into small pieces and placed in a Millipore filter eppendorf. After incubation at

-70°C for 1 hour to lyse agarose, DNA was spun through by centrifugation at 13000 rpm for 15 minutes.

There was less than 50% recovery of the PCR product by this method and hence it was necessary to use another method.

A2.1.6.3 Use of Mermaid kit

The PCR product was taken in 3 volumes of high salt binding solution (concentrated solution of sodium perchlorate). This mixture was then incubated at room temperature for 5 -15 minutes with GlassfogTM, a silica based matrix which binds the DNA. After centrifugation at 13000 rpm for 30 seconds to pellet GlassfogTM, the supernatent was kept aside. The pellet was washed with 200 µl of high salt binding buffer followed by a rinse with ethanol to remove salts and other contaminants. The pellet was then dried under vacuum to remove traces of ethanol. The pellet was resuspended in 16 µl water and incubated at room temperature for 5 minutes. DNA was eluted from GlassfogTM by centrifugation at 13000 rpm for 1 minute. The elution step was repeated and the DNA recovered was analysed on a 1% agarose gel. This method was found to be efficient for purifying low molecular weight DNA but not suitable for purifying the PCR product.

A2.1.6.4 Isolation of DNA fragments from agarose gel using DEAE membranes

This technique is derived from the method originally developed

independently by Winberg and Hammerskjold and Dretzen *et al* who used DEAE paper as the binding medium. The PCR product was allowed to run on a 1% agarose gel with TBE buffer containing ethidium bromide (0.4 μg/ml) until the tracking dye was in the centre of the gel. The gel was briefly visualised using a hand held transilluminator and the gel was slit horizontally 3 mm above and below the band with a sterile scalpel. Small pieces of DEAE cellulose paper were soaked in 2.5 M NaCl for several hours and subsequently washed with water and stored in 1 mM EDTA, pH 8 at 4°C. These pieces of DEAE cellulose were inserted into the two slits using a pair of tweezers and the gel was squeezed firmly against the paper to close the incisions. The paper below collects the PCR product and the paper above the band protects against contamination from high molecular weight DNA. Electrophoresis was resumed until all the DNA entered the lower paper (observed by examination under UV). The paper was washed in cold distilled water for 10 minutes, blotted dry with a tissue and placed in a Millipore filter eppendorf. The product was then placed in 0.5 ml extraction buffer (1.5M NaCl; 1mM EDTA; 20mM Tris HCl, pH 8) and incubated at 37°C for 2 hours after vigorous shaking to shred the paper. The mixture was then centrifuged at 13000 rpm for 5 minutes to separate the paper from the filtrate. DNA was extracted with 3 volumes of butanol (pre saturated with water) to remove ethidium bromide. The aqueous layer was then precipitated with ethanol and the pellet was resuspended in 32 μl of 10% TE, pH 8 buffer.

This method was found to recover greater than 50% of the PCR product and more efficient methods were sought.

A2.1.6.5 Purification of PCR product using Schleicher and Schuell NA-45 DEAE membrane

A modification of the DEAE paper method using Schleicher and Schuell NA-45 DEAE membrane was attempted to improve the efficiency of PCR product purification. DEAE cellulose in membrane form is a convenient medium for binding DNA and RNA from gels. DEAE membranes are superior to paper because of their high wet strength (wet membranes do not release fibres or particulates) and high matrix density (0.45 µM pore size). DNA and RNA binding is quantitative only upto 7 μgcm^{-2} while the maximum capacity is 20 μgcm^{-2}. To increase binding capacity, membranes were washed in 10 mM EDTA pH 7.6 for 10 minutes and 0.5 M NaOH for 5 minutes followed by a few rinses in distilled water. Membranes were stored at 4°C. The PCR product was then purified as described in method 5.4. The purity of the PCR product was adequate for further genetic manipulations including restriction enzyme digestions and cloning.

A 4 µl aliquot of this sample was run on a 1% agarose gel to check the size of the PCR product.

A2.1.6.6 Purification of PCR product using β-Agarase I enzyme

β-Agarase I can be used for the purification of both large (> than 50 kb) and small DNA fragments. It acts by cleaving carbohydrate bonds, freeing trapped DNA and producing carbohydrate molecules which can no longer gel.

The PCR product was allowed to run on a 2% low melting point agarose

gel at 4°C. The band running at about 600 base pairs was chopped and placed in a weighed eppendorf. The chopped band was weighed and then melted at 65°C for 15 minutes. The reaction mixture was heated at 65°C for 5 minutes in the presence of 10 mM Tris HCl, pH 6.5 and 1 mM EDTA. The reaction mixture was then placed at 40°C for 3 minutes prior to the addition of 4 units of beta- agarase enzyme (NEB). The reaction was carried out at 40°C for one hour. The enzyme was inactivated by heating at 85°C for 10 minutes and then immediately chilled on ice. After centrifugation at 13000 rpm for 30 minutes the supernatant was transferred into a fresh eppendorf. DNA was precipitated by the addition of one-tenth volume of 3 M sodium acetate, pH 5.2 and 2 volumes of isopropanol at -20°C. The supernatant was discarded after centrifugation at 13000 rpm for 30 minutes and the pellet resuspended in 12 µl water.

A 2 µl aliquot of this sample was run on a 1% agarose gel to check its purity and concentration.

A2.1.6.7 Purification of PCR products by electroelution

The PCR products were separated on a 1% low melting point agarose gel. The band running between 500 and 800 base pairs was chopped and placed in a dialysis tubing which was treated as described in Maniatis *et al* '82. 0.2 ml of TE buffer, pH 8 was added into the dialysis tubing. The chopped band was placed perpendicular to the applied current. The DNA migrates onto the walls of the tubing by applying a current of 60 mA and a voltage of 50 mV. The polarity was reversed for a minute to allow the DNA to move into the buffer from the walls of the tubing.

The DNA was precipitated by the addition of 1/10th volume of 3 M sodium acetate, pH 6 and 3 volumes of ethanol at -20°C overnight. After centrifugation at 13000 rpm for 15 minutes, the pellet was washed twice with 70% ethanol and dried. The pellet was then resuspended in 20 µl water. This method was found to be highly efficient with more than 90% recovery of purified products.

RACE products that appear as discrete bands and prove unambiguous can be cloned and sequenced directly, used as probes in Southerns, Northerns and in situ hybridisation analysis.

A2.1.7 Preparation of replicative form of M13mp18

A small scale culture of M13 infected cells was prepared by inoculating 50 ml of 1× TY with 50 µl of TG1 overnight and a 50 µl aliquot from a phage stock and grown at 37°C, 250 rpm overnight. The cells were pelleted by centrifugation at 4000 rpm for 15 minutes at 4°C. The pellet was resuspended in 100 µl STE buffer (0.2M NaCl; 10 mM Tris HCl, pH 8; 1 mM EDTA, pH 8). After centrifugation at 4000 rpm for 15 minutes at 4°C the pellet was resuspended in 2 ml of ice cold 50 mM glucose; 25 mM Tris HCl, pH 8; 10 mM EDTA (Solution 1) containing 10 mg/ml lysozyme and incubated at 4°C for 5 minutes. Cells were completely lysed by the addition of 4 ml of 0.2 M NaOH and 1% w/v SDS (Solution 2) and mixed gently on ice for 5 minutes. After the addition of 3ml of ice cold 5 M potassium acetate, pH 4.8 (Solution 3) the mixture was incubated on ice for 5 minutes. This mixture was then centrifuged at 10000 rpm at 4°C for 10 minutes. The supernatant

containing RF DNA was extracted once with phenol (water equilibrated): chloroform and the aqueous layer was precipitated with propanol by incubation at room temperature for 30 minutes followed by centrifugation at 10000 rpm for 10 minutes at 4°C. The pellet was resuspended in 4 ml ethanol. After centrifugation at 10000 rpm at 4°C for 10 minutes the pellet was resuspended in 1 ml TE buffer.

A2.1.8 Purification of M13mp18 by gel filtration on Sephacryl-300

The DNA pellet was resuspended in 6 ml of 0.1 M Tris HCl, pH 8; 0.01 M EDTA (TE buffer), transferred to a 30 ml Corex tube and 960 µl 2.5 M NaCl was added to the mixture. The solution was then precipitated overnight with 2.5 volumes ethanol after the addition of 5 µl of RNAse A (0.1 mg/ml). After incubation the solution was centrifuged at 13000 rpm for 10 minutes. The pellet was washed with 70% ethanol and resuspended in 500 µl water.

RF DNA was purified from RNA by gel filtration on a 10 ml column of Sephacryl-300 (Pharmacia) pre equilibrated with 0.2 M NaCl; 10 mM Tris HCl, pH 8; 0.5 mM EDTA. 20 fractions containing DNA were collected and 5 µl of each fraction was spotted onto 0.5% agarose containing ethidium bromide (10 mg/ml). After drying for 15 minutes at 37°C, the fractions containing DNA were identified with UV fluorescence. The first peak of fluorescence corresponds to RF DNA. The fractions were pooled and the DNA was precipitated by incubation with 0.5 volume of propanol for 30 minutes followed by centrifugation at 13000 rpm for 15 minutes at room temperature. The pellet was then resuspended in 50 µl TE buffer (pH 8).

A 4 µl aliquot of the sample was run on a 1% agarose gel along with M13mp18 single stranded DNA marker (Figure A2.5).

A2.1.9 Cloning of the PCR product into RF M13mp18

A Not I site was introduced into the AN Poly C primer while a Bgl II site was introduced into the C_β primer to make it convenient for cloning into M13mp18.

The purified PCR product was digested with Not I (10 U/µl) and Bgl II (10 U/µl) in the presence of buffer (10 mM Tris-HCl, pH 7.5; 7 mM $MgCl_2$; 100 mM NaCl and 7 mM 2-mercaptoethanol) by incubation at 37°C overnight. The insert was cleaned by extraction with an equal volume of phenol: chloroform followed by two extractions with chloroform. This product was then precipitated by the addition of 1/10th volume of 3M sodium acetate, pH 6 and three volumes of ethanol at -20°C. After centrifugation at 13000 rpm for 15 minutes, the pellet was rinsed once with 70% ethanol and dried. This cleaned insert was then cloned into an M13mp18 derived phage vector from which the M13mp18 gene II Bgl II site had been deleted by oligonucleotide site directed mutagenesis, and into which had been cloned an in frame poly linker fragment containing Not I and Bgl II restriction enzyme sites. This vector was a kind gift from Paul Moss from the Institute of Molecular Medicine, Oxford.

RF M13mp18 was similarly subjected to a double digestion with Not I and Bgl II restriction enzymes. 8 ng of the PCR product was ligated into 16 ng of Not I, Bgl II digested M13mp18 derived phage vector using 2.8 Weiss units of T4 DNA ligase in 30 mM Tris HCl, pH 7.4; 10 mM DTT; 10 mM $MgCl_2$; 10 mM ATP at

Figure A2.5 Single stranded template of TCR Vβ gene from KM vs Priess cells seen in lanes 2,3,5 and 6 along with marker seen in lane one.

16°C for 18hours.

A2.1.10 Transformation of TG1 cells

An aliquot (200 µl) of competent cells (as described in Appendix 3) was mixed with the ligation mixture in a pre cooled 15 ml polypropylene tube (Falcon). The mixture was incubated on ice for 1 hour before incubation at 42°C for 3 minutes (heat shocking). After 2 minutes on ice 200 µl of exponentially growing TG1 lawn cells, 40 µl IPTG (100mM), 40 µl 2% X-gal and 4 ml molten H-agar (at 42°C) were added. The mixture was immediately poured onto pre warmed 1 × TY plates and when set the plates were inverted and incubated at 37°C for at least 9 hours.

Single stranded DNA was prepared from the colourless plaques and screened for recombinants. Screening of recombinants can be done by a number of methods namely oligo probing, PCR screening, Southerns.

A2.1.11 PCR Screening

Protocol 1

After blue/white selection, white plaques were subjected to PCR screening. White plaques were inoculated into 41.5 µl water along with 200 nM concentration of AN PolyC and C_β primers, 200 µM dNTPs, 50 mM KCl, 10 mM Tris HCl, 0.1% Tween 20 and 1.5 mM $MgCl_2$. This mixture was denatured at 94°C for 5 minutes and 2.5 Units of Taq Polymerase (Perkin Elmer Cetus) was added.

The mixture was then overlaid with 2 drops of mineral oil and subjected to 25 cycles of denaturation at 94°C for 1 minute, annealing at 55°C for 1 minute and extension for 2 minutes at 72°C. It was finally subjected to an extension step at 72°C for 10 minutes.

Protocol 2

White plaques were inoculated into 10 µl water and denatured at 94°C for 10 minutes. The volume was made upto 50 µl in water with a final concentration of 200 nM AN Poly C and C_β primers, 200 µM dNTPs, 50 mM KCl, 10 mM Tris HCl, 0.1% Tween 20, 1.5 mM $MgCl_2$ and 2.5 units of Taq Polymerase (Perkin Elmer Cetus). It was then subjected to 25 cycles of denaturation at 94°C for 1 minute, annealing at 55°C for 1 minute and extension for 2 minutes at 72°C. It was then extended for 10 minutes at 72°C.

Protocol 3

RF from white and blue plaques was subjected to the same conditions of PCR screening as described in protocol 2. The volume was made upto 50 µl with water and subjected to 1 cycle of denaturation at 94°C for 5 minutes, annealing at 55°C for 2 minutes and extension at 72°C for 10 minutes. It was then subjected to 30 cycles of denaturation at 94°C for 45 seconds, annealing at 55°C for 2 minutes and extension at 72°C for 10 minutes.

A2.1.12 Oligo probing

A2.1.12.1 Design of the probe

A probe RADS (Table A2.2) was designed by checking against the DNA sequences of human TCR β chain C region exon 1 using an identity search program. The oligo was synthesised on an Applied Biosystems DNA Synthesiser and purified as described in Section A2.1.4.1 using phosphoramidate chemistry.

A2.1.12.2 5' phosphorylation of the oligo probe

RADS probe was 5' phosphorylated. A 48 picomole sample of probe was mixed with buffer (0.05 M Tris HCl, pH 7.6; 0.01 M $MgCl_2$; 5 mM DTT; 0.1 mM Spermidine HCl; 0.1 mM EDTA, pH 8) 100 µCi $\gamma^{32}P$ ATP and 7 units T4 polynucleotide kinase enzyme in a final volume of 20 µl. The reaction was carried out at 37°C for 1 hour and stopped by the addition of 20 mM EDTA, pH 8. The phosphorylated probe was purified on a P-30 column to remove non incorporated ATP. This probe was loaded onto a 1 ml P-30 column pre equilibrated with 0.2 M NaCl; 10mM Tris HCl, pH 8; 1 mM EDTA, pH 8. A 100 µl fraction containing the purified probe was collected after centrifugation at 1600 rpm for 5 minutes. The concentration of the purified probe was determined spectrophotometrically.

A2.1.12.3 Determination of efficiency of the phosphorylation reaction

The efficiency of the phosphorylation reaction was determined by measuring the proportion of radioactive precursors incorporated into the probe. This was done by differential adsorption of the product onto a positively charged surface like DE-81 (as described in Section A2.1.2.2).

1 µl of the reaction volume was spotted onto the centre of 2 DE-81 filters. One is used to measure the total amount of radioactivity in the reaction. This filter was washed thrice with 0.5 M disodium hydrogen phosphate, pH 7 for 2 minutes. It was then washed in water for 1 minute followed by a wash in 70% ethanol for 4 minutes and then dried. Each of the filters (washed and unwashed) was inserted into scintillation vials and the amount of radioactivity on each filter was measured. ^{32}P can be determined on dry filters by Cerenkov counting (in the ^{3}H channel of a liquid scintillation counter). The amount of radioactivity on the unwashed filter can be compared with the amount of radioactivity on the washed filter and the % incorporation calculated. Only probes with a specific activity of 5×10^5 to 2 X 10^6 cpm/ml were used for the hybridisations.

A2.1.12.4 Plaque lift

Protocol 1

Recombinant plaques were spotted onto a grid and incubated at 37°C overnight, to reduce background signals. These plaques were then transferred onto

a nylon membrane (Hybond-N) by diffusion at room temperature. The master plate was stored at 4°C.

This membrane was placed on a Whatmann 3MM filter soaked in 10% SDS for 5 minutes. DNA was denatured by incubation in 1.5 M NaCl; 0.5 M NaOH for 5 minutes and then neutralised on Whatmann 3 MM filter soaked in 1 M Tris HCl, pH 7.5; 1.5 M NaCl for 10 minutes. The membrane was then dried and baked at 80°C for 2 hours. It was washed in 50 mM Tris HCl, pH 8; 1M NaCl; 1.25 mM EDTA, pH 8; 0.1% SDS to remove bacterial debris.

Membranes were prehybridised in 5 X SSC; 5 X Denhardts solution; 50 µg/ml sonicated salmon sperm DNA; 50 mM NaPi pH 6.8; 1 mM NaPPi; 100 µM ATP and 20% formamide at 42°C for 12 hours and hybridised with 1 X 10^6 cpm/ml of labelled probe (as described in section 12.1.2) at 42°C for 16 hours. For low stringency washes, membranes were washed in 2 X SSC; 0.1% SDS at 40°C for 20 minutes four times. Stringency can be increased by raising the temperature and lowering the salt concentration. Membranes were dried prior to being exposed overnight to X-Ray film at -70°C.

Genuine signals resulting from hybridisation of the probe to cloned DNA can be distinguished from background artifacts by probing two filters from the same master plate. Only signals that duplicated were regarded as authentic positives.

The phage plaques giving positive signals were purified and single stranded DNA was prepared from them. These recombinants were sequenced by Sanger's dideoxy chain termination method (as described in Appendix 3).

Protocol 2

Recombinant plaques were transferred onto a nylon membrane (Hybond-N), cut to the size of the plate, by diffusion at room temperature. The membrane was placed over the plaques for 2 minutes. The master plate was stored at 4°C. DNA was denatured by placing the membrane on a Whatmann 3MM paper soaked in 0.5M NaOH for 3 minutes. DNA was neutralised by washing the membrane twice in 1M Tris HCl pH 7.5 for a minute each, followed by a wash in 1M Tris HCl pH 7.5, 1.5M NaCl for 5 minutes.

The membrane was dried and baked at 80°C for 2 hours to immobilise the DNA. It was then washed in 100ml 6 X SSC at 67°C for 5 minutes followed by a wash in 10 X Denhardts, 6 X SSC, 0.2% SDS at 67°C for 5 minutes and finally a rinse with 100ml 6 X SSC for 5 minutes. The C_β specific probe RADS was radiolabelled (as described in section 2.1.12.2), added to 8mls 6 X SSC and stored at -20°C till further use. Membranes were hybridised with 1 X 10^6 cpm/ml of labelled probe at 67°C for 30 minutes. It was then cooled to room temperature. The membrane was washed in 100ml 6 X SSC thrice. After draining off the SSC solution it was covered with clingfilm and exposed overnight to X-Ray film at -70°C. The membrane was then washed at 65°C for 1 minute in 6 X SSC twice and exposed to film. Similarly it was washed at 75°C for 1 minute in 6 X SSC and exposed to film at -70°C (Figure A2.6). The phage plaques giving positive signals were purified and single stranded DNA was prepared from them. These recombinants were sequenced by dideoxy chain termination method as described in Appendix 3.

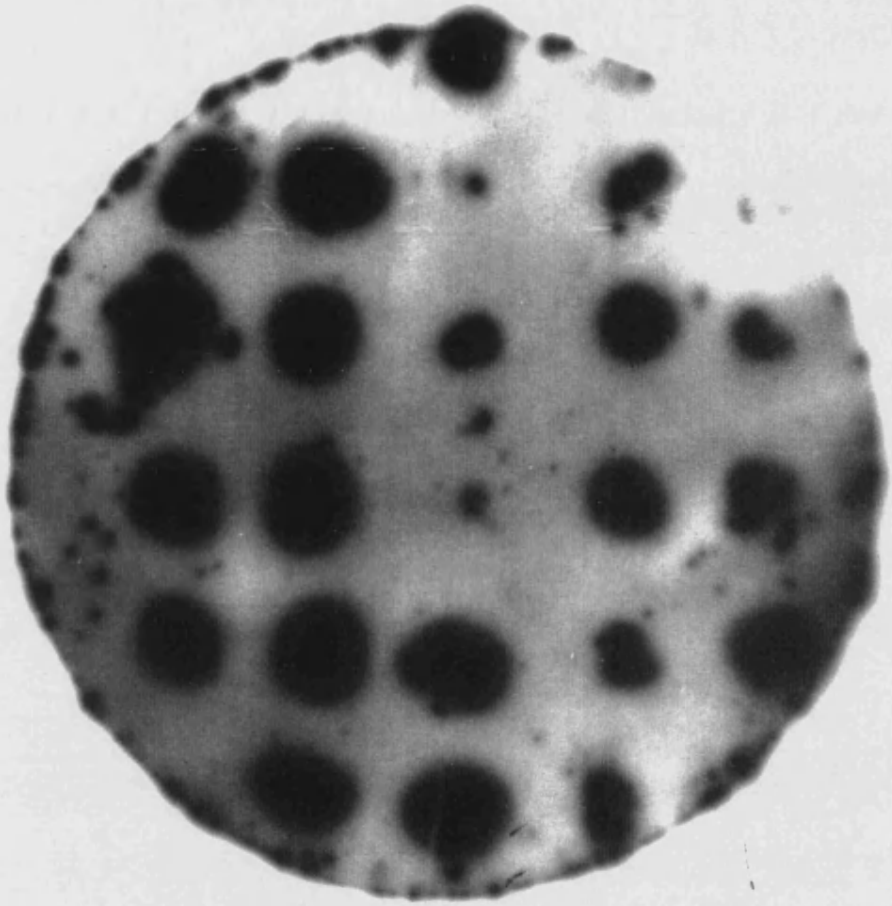


Figure A2.6 Oligo probing of the TCR Vβ gene obtained from KM vs Priess cells.

Only the recombinant signals which duplicated were sequenced.

A quicker strategy to obtain sequence data from PCR products was devised. This strategy bypasses the time consuming cloning steps. The PCR product is purified by electro elution and then subjected to Southern blotting. All positives are directly sequenced by PCR sequencing.

A2.1.13 Transfer of DNA onto nylon filters by Southern blotting

RACE products can be examined by Southern blotting to confirm the presence of desired cDNAs. The purified PCR product was subjected to agarose gel electrophoresis. The gel was immersed in 0.25M HCl for depurination of DNA for 7 minutes. The gel was then soaked in denaturing solution (0.5M NaOH, 1.5M NaCl) for 30 minutes at room temperature. When the denaturation was complete the bromophenol blue dye turned light blue. The gel was then neutralised in 0.5 M Tris HCl pH 7.5, 3M NaCl for 30 minutes after a rinse with distilled water. The gel was then placed on 6 sheets of pre wetted (20 X SSC) Whatmann 3MM paper. A pre wetted (10 X SSC) nylon filter was placed on the gel. Four dry Whatmann 3 MM sheets were placed on a pre wetted Whatmann 3 MM sheet. All the air trapped in the filter was removed and the filter was wetted again. A cling film was placed on the stack in such a way as to prevent contact between the reservoir and the 15 cm paper towels kept above the basic sandwich placed in a reservoir of 20 X SSC. A glass plate was placed on top of this stack along with a 500 gm weight. The transfer was allowed to take place overnight

by capillary blotting. The filter was then air dried. Transferred DNA was then fixed by baking in an oven at 80°C. Alternatively DNA can be fixed using a U.V. cross linker for 5 minutes. Filters can be stored in the dark at room temperature for a long period.

The filter was pre hybridised in 3 X SSC, 2 X Denhardts solution, 20 µg/ml salmon sperm DNA, 6% PEG 6000) at 65°C for 1 hour. The filter was then placed in 100 ml hybridisation solution (5 X Denhardts solution, 3 X SSC, 20 µg/ml salmon sperm DNA, 6% PEG 6000) at 65°C overnight with a probe having a specific activity between 5 X 10^5 and 2 X 10^6 cpm/ml. Probes were prepared as described in section A2.1.12.2.

Low stringency washes were carried out in 2 X SSC at 45°C three times. For high stringency, filters were washed twice in 2 X SSC, 0.1 % SDS at 65°C for 15 minutes followed by two washes in 0.1 X SSC, 0.1 % SDS at 65°C for 15 minutes. The filter was then air dried, wrapped in cling film and subjected to autoradiography at -70°C (Figure A2.7).

A2.1.14 PCR Sequencing

This technique also known as linear amplification sequencing is a variation of PCR. The components of a chain termination sequencing reaction are cycled through a temperature profile consisting of a heat denaturation step, an annealing step and an extension step. There are three basic differences between cycle sequencing and PCR amplification. Only one primer is necessary in

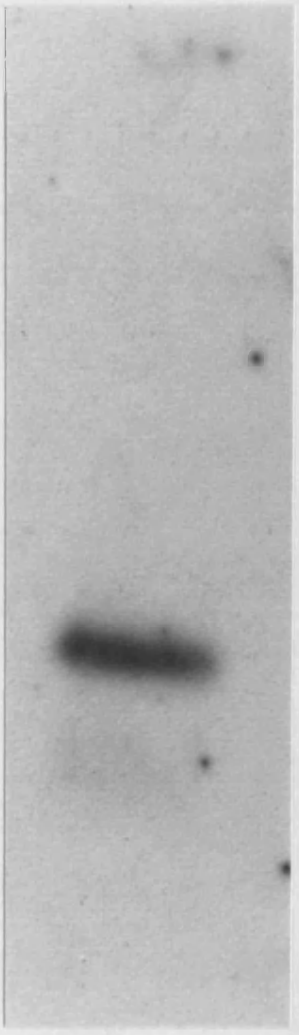


Figure A2.7 Southern blot of TCR Vβ gene from KM vs Priess cells.

the cycle sequencing reaction to prime synthesize a strand of DNA. The resultant linear amplification of the sequencing product leads to an increase in the signal generated during the sequencing reaction. Another difference is the presence of dideoxy nucleotide triphosphates in the sequencing reaction which create the base specific terminations required for sequencing. The last difference is the addition of an α-labeled nucleotide triphosphate to be incorporated into the growing DNA strand, providing the ability to detect the sequence ladder by autoradiography.

Protocol 1

90 pm of RADS primer was end labelled with 10 μCi of ^{32}P (ATP) in the presence of 0.05 M Tris HCl, pH 7.6; 0.01 M $MgCl_2$; 5 mM DTT; 0.1 mM spermidine HCl; 0.1 mM EDTA, pH 8 and 7 units of T4 Polynucleotide kinase. The reaction was carried out at 37°C for 30 minutes. The reaction was terminated by heating the mixture to 90°C for 2 minutes to inactivate the kinase. Pure PCR product was taken in a reaction mixture with 22.5 pm of RADS primer; 2.5 units of Taq Polymerase; 50 mM KCl; 10 mM Tris HCl; 0.1% Tween 20 and 1.5 mM $MgCl_2$.

To 4 μl of this annealed primer-template 4 μl of dideoxy termination mixtures was added. This mixture was overlayed with a drop of mineral oil. Samples were taken in duplicate. One set was subjected to 10 cycles of denaturation at 95°C for 40 seconds, annealing at 45°C for 30 seconds and extension at 72°C for 2 minutes. The other set was subjected to 10 cycles of denaturation at 95°C for 40 seconds, annealing at 52°C for 30 seconds and extension at 72°C for 2 minutes.

This reaction was stopped by the addition of 4 µl of formamide dye. After denaturation at 90°C for 2 minutes, 3 µl of this sample was loaded onto an 8% polyacrylamide gel.

Protocol 2

This technique is based on the ability of Taq Polymerase to extend a labelled primer hybridised to the template to be sequenced until a chain terminating nucleotide is incorporated.

For 5 reactions, 10 pmoles of C_{reg} primer (66 ng for 20mer) was kinased with 10 pmoles of $\gamma^{32}P$ (dATP)in the presence of 0.05 M Tris HCl, pH 7.6; 0.01 M $MgCl_2$; 5 mM DTT; 0.1 mM spermidine HCl; 0.1 mM EDTA, pH 8 and 7 units T4 Polynucleotide kinase. The reaction mixture was incubated at 37°C for 15 minutes. The reaction was terminated by heating the mixture at 90°C for 2 minutes. The efficiency of labelling was determined by finding the % incorporation as described in section 2.2.2. The purified PCR product was taken with 2 pmoles of labelled primer in the presence of 50 mM KCl, 10 mM Tris, 1.5 mM $MgCl_2$, 0.1% Tween 20 and 2.5 units Taq Polymerase. The volume was made upto 17.5 µl with water. 4 µl of labelled primer/template was taken separately with 4 µl ddATP, ddCTP, ddGTP and ddTTP terminating mixture in 4 tubes marked "A", "G", "C" and "T". These mixtures were overlayed with 2 drops of mineral oil and then subjected to 10 cycles of denaturation at 94°C for 40 seconds, annealing at 55°C for 30 seconds and extension at 72°C for 2 minutes. The reaction was stopped by the addition of 4 µl formamide dye. 3 µl aliquot of this mixture was run on an 8% polyacrylamide gel after denaturation at 75°C for 3 minutes.

Protocol 3

18 pm of C_{reg} primer was end labelled with 10 µCi of ^{32}P (ATP)in the presence of 0.05 M Tris HCl, pH 7.6; 0.01 M $MgCl_2$; 5 mM DTT; 0.1 mM spermidine HCl; 0.1 mM EDTA pH 8 and 7 units of Polynucleotidyl kinase. This reaction was incubated at 37°C for 30 minutes and terminated by heating the mixture at 55°C for 5 minutes. The labelled primer was taken in the presence of 50 mM KCl; 10 mM Tris HCl; 0.1% Tween 20; 1.5 mM $MgCl_2$; template DNA and 2.5 units of Taq Polymerase. The volume of the pre reaction mixture was made upto 36 µl with water. 8 µl of this pre reaction mixture was taken along with 2 µl of ddA, ddG, ddC and ddT termination mixtures respectively. After overlaying this mixture with one drop of mineral oil it was heated at 95°C for 5 minutes. In the case of partially pure PCR products, the reaction mixture was subjected to 20 cycles consisting of a denaturation step at 95°C for 30 seconds, annealing step at 55°C for 30 seconds and extension at 72°C for 1 minute followed by 10 cycles consisting of a denaturation step at 95°C for 30 seconds and an extension step at 70°C for 1 minute. Pure PCR products were subjected to 20 cycles of denaturation at 95°C for 5 seconds and annealing at 55°C for 5 seconds. After the addition of 5 µl of the stop solution (95% formamide; 20 mM EDTA; 0.05% Bromophenol Blue; 0.05% Xylene Cyanol FF) the mixture was heated to 90°C for 5 minutes. A 3 µl aliquot was loaded onto an 8% polyacrylamide gel.

Protocol 4

Pure PCR product was taken in the presence of 40 mM Tris Hcl, pH 7.5; 20

mM $MgCl_2$; 50 mM NaCl and 18 pm of AN Poly C primer. This template-primer mixture was boiled for 5 minutes, mixed and flash frozen in dry ice/ ethanol. The reaction mixture was then allowed to anneal at room temperature for 30 minutes. The sequencing reactions were then carried out as recommended in the DNA Sequencing kit supplied by USB. This method was used subsequently for direct sequencing of PCR products.

A2.1.15 FASTA Analysis

Several GCG programs like FASTA and BLAST.can be used to scan sequence databases. BLAST or Basic Local Alignment Search Tool, uses the method of Altschul *et al* (1990) to search for similarities between the query sequence and all the sequences in a sequence database. It is a statistically driven search method that finds regions of similarity between the query sequence and the database. These regions of similarity are known as segment pairs. It can either search databases on a personal computer or databases maintained at the National Centre for Biotechnology Information (NCBI).

BLAST is not as sensitive as FASTA in searching nucleotide sequence databases as it ignores high scoring segment pairs that do not contain a perfect match of at least 12 nucleotides (24 bits). Many obviously significant relationships are not found although nearly identical sequences are found. Since FastA finds similarities that BLAST has missed, it was used to determine the identity of the unknown sequence.

FASTA uses the method of Pearson and Lipman (1988) to search for

similarities between the query sequence and any group of sequences. This method first identifies the ten best regions of similarity between the query sequence and each sequence from the search set using a modification of the algorithm of Wilbur and Lipman (1983). These searches belong to a class of comparisons that use 'direct addressing' or 'k-tuple preprocessing' to increase efficiency.

In the first step of this search the comparison can be viewed as a set of dotplots with the query as the vertical sequence and the group of sequences to which the query is being compared as the different horizontal sequence. First step finds the registers of comparison (diagonals) having the largest number of short perfect matches (words) for each comparison. In the second step, these 'best' regions are rescored using a scoring matrix that allows conservative replacements, ambiguity symbols and runs of identities shorter than the size of a word. In the third step, the program checks to see if some of these initial highest scoring diagonals can be joined together. The highest scoring search set sequences are aligned to the query sequence for display. The best scores are displayed below the histogram. It displays the alignments of the regions of best overlap between the query and search sequences. The scores for matching the new sequence against each data base protein was noted and the highest scoring matches are examined as potential homologies

FASTA analysis of the sequence obtained from the KM vs Priess cell line established its identity as a TCR Vβ sequence. The database sequence which showed the highest homology is displayed in Figure A2.8. Thus the strategy outlined above can be successfully used to obtain TCR Vβ gene sequence information from T cells.

```
(Nucleotide) FASTA of: rads.seq  from: 1 to: 250  January 11, 1994  13:45

beta tcr seq

 TO: nucleic:*  Sequences:      3,355  Symbols:  8,128,496  Word Size: 6


The best scores are:                                          initl initn opt..

nucleic:rwhuby /rev  T-cell receptor beta chain V and C r... 306   306   422
nucleic:tvmvtc /rev  Transduced T-cell antigen receptor b... 225   273   281
nucleic:rwmsb /rev  T-cell receptor beta chain precursor ... 239   239   261
nucleic:rwms86 /rev  T-cell receptor beta chain V and C r... 203   231   273
nucleic:rwms20 /rev  T-cell receptor beta chain V and C r... 179   207   255
nucleic:rwrbb /rev  T-cell receptor beta chain mRNA (ANA ... 193   193   201
nucleic:rwmsb2 /rev  T-cell receptor beta-2 chain C regio... 175   175   178
nucleic:rwmsb1 /rev  T-cell receptor beta-1 chain C regio... 175   175   178

rads.seq /rev
nucleic:rwhuby

R;Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S.P., Aleksander, I., and
Mak,
 T.W.
Nature 308, 145-149, 1984 (Residues 1-1151)
F;38-973/Protein: T-cell receptor beta chain YT35

SCORES     Initl: 306 Initn: 306 Opt: 422
            58.9% identity in 219 bp overlap

              10         20         30         40         50
rads.s CAGTGATCGCTCTTGCAGAGCTGAGGATCGTCTCCACTCTGAAGATCCAGCGCACACA--
                                    :|:||||:|:|||||:||||| | || |
rwhuby GAUUCUCAGCUAAGAUGCCUAAUGCAUCAUUCUCCACUCUGAAGAUCCAGCCCUCAGAAC
          290        300        310        320        330        340

      60         70         80         90        100        110
rads.s GCAGGAGACTCCGCCGTG-ATCTCTGTGCCAGCAGCTCCGGGTTTCTAGCGGGAGCT-AC
        |||| |||:| || |:| |  :|:|:|||||||| : |  |        ||   ||: ||
rwhuby CCAGG-GACUCAGCUGUGUACUUCUGUGCCAGCAGUUUCUCGACCUGUUCG---GCUAAC
           350        360        370        380        390          400

       120        130        140        150        160        170
rads.s AATGAGCAGTTCTTCGGGCCAGGGACACGGCTCACCGTGCTAGAGGACCTGAAAAACGTG
        |:|   |   |::|||  | |||||  || : ||||:  :|||||||||:||| || |:|
rwhuby UAUGGCUACACCUUCGGUUCGGGGACCAGGUUAACCGUUGUAGAGGACCUGAACAAGGUG
              410        420        430        440        450        460

       180        190        200        210        220        230
rads.s TTCCCACCCGAGGTCGCTGTGTTTGAGCCATCAGAAGCAGAGATCTAGCTTATCGATACC
       ::|||||||||||:|||:|:|:::|||||||:|||||||||||||:|:  |  | | | |
rwhuby UUCCCACCCGAGGUCGCUGUGUUUGAGCCAUCAGAAGCAGAGAUCUCCCACACCCAAAAG
              470        480        490        500        510        520

       240        250
rads.s GTCGACCTGCAGG
       | |   |:|
rwhuby GCCACACUGGUGU
              530
```

Figure A2.8 FASTA analysis shows the high homology displayed by the sequence obtained from KM vs Priess cell line (rads) to other database TCR sequences.

A2.2 V_α gene sequence analysis

V_α gene was purified from the T cell line KM vs Priess using the method described below.

A2.2.1 RACE method for V_α gene amplification

A2.2.1.1 Design, synthesis and purification of C_α primer

C_α 3' primer (Table A2.2) was designed for use in V_α Anchor PCR. A Sal I site was introduced into the primer to facilitate cloning into M13mp18. This primer was synthesised on an Applied Biosystems DNA synthesiser using phosphoramidate chemistry. The oligo sequences were checked against the DNA sequences of human TCR chain constant region exon 1 using an identity search program. Ammonia cleaves the oligo from the resin at room temperature. The protecting groups were removed from the terminus by incubating the resin in 0.88N ammonia solution at 55°C for 16 hours. Ammonia was removed by drying the oligo under vacuum. Ethanol precipitated oligo can be recovered by a 10 minute spin at 10000 rpm followed by a rinse with 70% ethanol and then dried. Concentration of the oligo was determined spectrophotometrically.

A2.2.1.2 5' phosphorylation of oligo primers

The C_α 3' primer was 5' phosphorylated. A 2 picomole sample of primer was mixed with buffer (0.05 M Tris HCl, pH 7.6; 0.01 M $MgCl_2$; 5 mM DTT; 0.1 mM spermidine HCl ; 0.1 mM EDTA, pH 8) and 1 unit T4 PNK enzyme in a final volume of 20 µl. The reaction was performed at 37°C for 30 minutes and

the kinase was inactivated by heating at 70°C for 10 minutes. The 5' phosphorylated primers were stored at -20°C.

A2.2.1.3 Preparation of template for Anchor PCR

Total RNA was extracted from a DR4 Dw4 restricted T cell clone HF53 by the RNAzol method (as described in Section 2.1.3). 1 µg of total RNA was used for cDNA synthesis using 0.5 µg of Oligo $(dT)_{12-18}$ annealed at 70°C for 10 minutes. Reactions were performed using 0.5 mM dNTPs; 10 mM DTT; 75 mM KCl; 50 mM Tris HCl (pH 8.3); 3 mM $MgCl_2$; 10 µCi of α-^{32}P dCTP and 200 units of Moloney Murine Leukemia Virus Reverse Transcriptase (BRL) at 37°C for 60 minutes. The reaction was terminated by the addition of 4 µl of 0.5 M EDTA, pH 8. The quality of cDNA was checked by running a 4 µl aliquot of the sample on a 1% alkaline gel.

The cDNA was made upto 100 µl and precipitated twice with 4 M ammonium acetate and three volumes of ethanol. The precipitate was taken up in 500 µl 70% ethanol and spun for 15 minutes. The pellet was then resuspended in 37.775 µl water and dG tailed using 0.5 M potassium cacodylate (pH 7.2); 10mM $CoCl_2$; 1mM DTT and 24.5 µM dGTP for 30 minutes at 37°C. The tailed cDNA was then diluted to a volume of 100 µl. It was extracted with an equal volume of phenol:chloroform mixture and then twice with chloroform. The aqueous phase was used as template for A-PCR.

A2.2.1.4 Anchor PCR

8 µl of the tailed cDNA was subjected to Anchor PCR using 2.5 units of Taq Polymerase (Perkin Elmer) and a Perkin Elmer Cetus Thermocycler under the following conditions: 94°C, 5 minutes; 55°C, 2 minutes; 72°C, 10 minutes for one cycle followed by denaturation at 94°C for 45 seconds, annealing at 55°C for 2 minutes and extension at 72°C for 10 minutes for 30 cycles. The reaction was performed using 200 µM dNTPs, 50 mM KCl, 10 mM Tris HCl, 0.1% Tween 20 and 1.5 mM $MgCl_2$. The primers used for V gene amplification were AN Poly C and C_α 3' at a concentration of 8.9 pm. The volume was made upto 50 µl.

A 5 µl aliquot of each sample was run on a 1% agarose gel along with ϕX 174 RF DNA marker (Figure A2.9). These PCR products were then purified by electroelution.

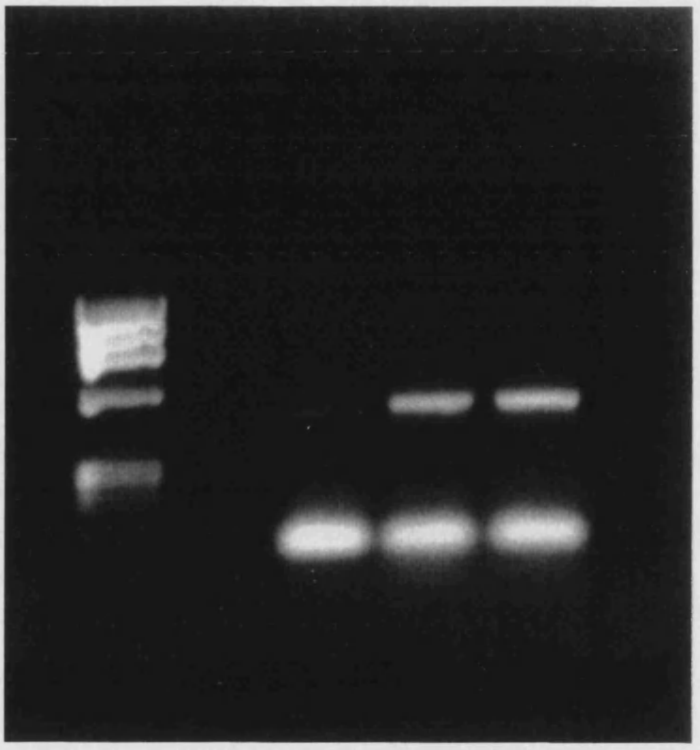


Figure A2.9 Anchored PCR product of TCR Vα gene from KM vs Priess cells seen in lanes 4 and 5 along with ϕX 174 RF DNA marker in lane 1.

APPENDIX-3

Standard Biochemical and Microbiological Techniques

A3.1 *Materials*

A3.1.1 Chemical reagents

Chemical reagents (analytical grade) were purchased from Sigma, Fisons and BDH except where stated otherwise.

A3.1.2 Restriction enzymes and DNA modifying enzymes

All enzymes were purchased from Pharmacia except where stated otherwise.

A3.1.3 Radiochemicals

All radiochemicals were purchased from NEN.

A3.1.4 Media and Antibiotics

Agar (No. 1), tryptone and yeast extract were purchased from Difco and antibiotics from Sigma. All media and antibiotics stock solutions were prepared as described in Sambrook *et al*,'89.

A3.2 Preparation of solutions

A3.2.1 RNase free Dnase

Pancreatic RNase A was dissolved in 50 mM EDTA to 10 mg/ml and contaminating DNase free was heat inactivated by incubating at 70°C for 15 minutes. The solution was aliquoted and stored at -20°C.

A3.3 Bacterial strains, plasmids and phage

E.coli strains BMH71.18 and W3110 were used for Gloop2DDFv protein expression while TG1 was used for all cloning purposes. The expression vector used was pUC19. M13mp18 derived phage vector was used for TCR cloning.

A3.4 Maintenance of bacterial strains, plasmids and phage

A3.4.1 Bacterial strains and plasmids

Stocks of bacterial strains and plasmids in the appropriate host cell were prepared for long term storage at -70°C by mixing 0.9 ml of a fresh overnight culture with 0.6 ml of sterile 80% glycerol. For short term storage, strains were streaked out onto agar or glucose minimal media plates, sealed with Nescofilm and stored at 4°C.

A3.4.2 M13mp18 phage stocks

A 5 ml aliquot of 1 × TY was inoculated with a single plaque and 500 µl

of a 1: 100 fold dilution of an overnight culture of TG1 was added. The culture was incubated with rolling at 37°C for 5 hours. 1.5 ml aliquots were dispensed into sterile microfuge tubes and centrifuged for five minutes to remove bacterial cells. One ml of the supernatant was transferred to a fresh sterile tube which was sealed with nescofilm and stored at 4°C.

A3.5 Isolation of plasmid from E.coli

A3.5.1 Small scale preparation of plasmid DNA

The rapid alkaline lysis method was used to isolate plasmid DNA from 1.5 ml of overnight culture. An isolated colony was picked and used to inoculate 10 mls of liquid media. After overnight growth at 37°C, 250 rpm, a 1.5 ml aliquot of the culture was transferred to a microfuge tube and centrifuged for 5 minutes. The pellet cells were resuspended in 100 µl of ice cold 50mM glucose; 25mM Tris-HCl, pH 8.0, 10mM EDTA (Solution I) containing 2 mg/ml lysozyme and incubated at 4°C for 5 minutes. Cells were completely lysed by the addition of 200 µl of 0.2M NaOH and 1% w/v SDS (Solution II) and mixed gently on ice for 5 minutes. 150 µl of ice cold 5M potassium acetate, pH 4.8 [Solution III: This solution is 3 M with respect to sodium and 5 M with respect to potassium. (To 29 ml of glacial acetic acid, on ice, add water to 80 ml; adjust the pH to 4.8 with 10 M KOH and then add water to a final volume of 110 ml) was added and the mixture incubated on ice for 5 minutes. Chromosomal DNA and cellular debris were pelleted by centrifugation at 13000 rpm for 5 minutes. The supernatant containing the plasmid DNA was extracted once with an equal volume of phenol

(water equilibrated): $CHCl_3$ and twice with ether (water equilibrated). DNA was precipitated from the aqueous solution by the addition of 0.1 volume of 3M sodium acetate (pH 5.2) and 3 volumes of ethanol at -20°C. After centrifugation at 13000 rpm for 15 minutes, the pellet was washed in 70% ethanol to remove the salt residue and dried. The pellet was resuspended in 50 µl water and stored at -20°C.

A3.5.2 Large scale preparation of plasmid

500 ml culture was prepared by inoculating media with 1ml of an overnight culture grown at 37°C, 250 rpm. As in the case of small scale preparations the DNA was isolated by the method of Birnhoim and Doly (1979) with modifications. The cells were pelleted by centrifugation using a GSA rotor in a Sorval RC-5 centrifuge by accelerating the rotor to 10,000 rpm. The cells were resuspended in 2 mls of Solution I, transferred to a 15 ml Corex tube and lysozyme added at a concentration of 5 mg/ml. The solution was incubated at 37°C for 20 minutes and then 4 ml of Solution II was added and incubated at 4°C. 3 mls of Solution III was added and the solution incubated at 4°C for 5 minutes. After 10 minutes centrifugation at 10,000 rpm the supernatant was poured off and extracted once with an equivalent volume of Tris buffered Phenol:Chloroform. The DNA was precipitated by the addition of 0.6 volume of isopropanol and incubated at room temperature for 30 minutes. After centrifugation, the pellet containing plasmid DNA and contaminating RNA was washed in 100% ethanol to remove all traces of isopropanol and left to dry. The DNA was then purified from RNA by caesium

chloride gradient purification.

A3.5.3 Caesium chloride gradient purification

The dried DNA pellet was resuspended in 1 ml of TE buffer and the volume made up accurately to 2.4 mls with TE buffer .Then .4.2 g caesium chloride was dissolved into the DNA solution and 400 µl of ethidium bromide (10 mg/ml) was added. 8 mls of a salt cushion, made by dissolving 15.41 g of caesium chloride into 20 ml of TE buffer was transferred into a Beckman Quick Seal ultracentrifuge tube, with care being taken not to introduce air bubbles. The DNA/Ethidium bromide solution was then carefully introduced into the centrifuge tube below the salt cushion using a syringe. The tube was filled to the neck with the cushion solution and heat sealed. Caesium chloride gradient was established by centrifugation in a Beckman (8-M ultracentrifuge using a [70.1Ti] rotor) accelerating to 55,000 rpm for 17 hours. The DNA bands were identified by ultraviolet fluorescence and the solution removed with a syringe. The contaminating RNA and proteins were pelleted at the bottom of the tube. The DNA solution was extracted 3 times with water- equilibrated butanol to remove the ethidium bromide and then an equal volume of TE buffer was added to dilute the caesium chloride. The DNA was precipitated in 0.6 volumes of propan-2-ol for 30 minutes at room temperature and the pellet resuspended in 300 µl 0.3M sodium acetate pH 5.5. After transferring to a microfuge tube, the DNA was reprecipitated with ethanol. The pellet was washed in 70% ethanol, dried and resuspended in 100 µl water.

A3.5.4 Preparation of high quality plasmid DNA without caesium chloride gradient.

100 ml 2 X TY was inoculated with 1 ml of an exponential culture (grown at 37°C, 250 rpm till an optical density of 0.3-0.5) and grown at 37°C, 250 rpm overnight. The culture was pelleted by spinning twice at 5000 rpm for 10 minutes. The pellet was resuspended in 4 ml of TEG (25 mM Tris HCl, pH 8.0; 50 mM glucose; 10 mM EDTA) and the solution was vortexed to get rid of clumps. 8 mls of lysis solution (0.2 M NaOH, 1% SDS) was added till the solution became viscous. This solution was then left on ice till it became transparent. After the addition of 6 mls of ice cold potassium acetate solution, it was kept on ice for 15-30 minutes. The solution was then spun at 5000 rpm for 15 minutes. 18 mls propan-2-ol was added to the supernatant and the mixture was kept on ice for 15 minutes. It was then spun at 5000 rpm for 30 minutes. The supernatant was discarded and the inside of the tube was wiped with tissue. The pellet was then resuspended in 2 mls of TE buffer. After the addition of 2 mls of 6M LiCl solution it was kept on ice for 15 minutes. RNA and protein were precipitated by centrifugation at 5000 rpm for 10 minutes. The supernatant was then transferred to another tube and after the addition of 10 mls of absolute ethanol it was left on ice for 30 minutes. After centrifugation at 5000 rpm for 30 minutes the supernatant was discarded and the pellet resuspended completely in 2 mls 70% ethanol to form a turbid solution. It was then respun at 5000 rpm for 15 minutes. The pellet was resuspended in 500 µl TE buffer. After the addition of 1 ml of absolute ethanol the solution was allowed to stand on ice for 15 minutes. 50 µl 3 M sodium acetate, pH 6 was added to the

mixture at this stage if no precipitate was observed. It was then incubated on ice for 15 minutes. After centrifugation at 13000 rpm for 30 minutes the pellet was rinsed with 70% ethanol. It was then respun at 13000 rpm for 5 minutes. The pellet was dried for 5 minutes under vacuum and then resuspended in 400 µl TE buffer. It was incubated at 37°C for 30 minutes after the addition of 4 µl of Bovine Pancreatic RNase fraction V (10mg/ml). After the addition of 20 µl 10% SDS it was heated at 75°C for 10 minutes. It was kept at room temperature for 15 minutes after the addition of 400 µl 6 M LiCl solution. After centrifugation at 13000 rpm for 10 minutes the supernatant was taken in 2 eppendorf tubes. It was kept on ice for 30 minutes after the addition of 1 ml absolute ethanol. If no precipitate was visible at this stage, 40 µl 3M sodium acetate, pH 6 was added and the mixture was left on ice for another 30 minutes. After centrifugation at 13000 rpm for 30 minutes the pellet was rinsed with 70% ethanol. The pellet was resuspended in 400 µl water after centrifugation at 13000 rpm for 10 minutes. After one extraction with phenol/chloroform/isoamyl alcohol it was extracted twice with chloroform. The DNA was precipitated with ethanol and resuspended in 100 µl water.

A3.6 Agarose gel electrophoresis of nucleic acids

Agarose gels were cast and run in a Bio-Rad Mini-gel apparatus to form gels of dimensions 10cm × 6cm × 0.5cm. Native gels contained varying concentrations of agarose (0.8 -1.0% [w/v]) and ethidium bromide at a concentration of 0.4 µg/ml. The DNA samples were mixed with 0.1 volume of 10 × loading buffer (3.5% w/v Ficoll; 0.1% w/v bromophenol blue; 0.1% w/v xylene

cyanol and 4mM EDTA) and the gels were run in 1 × TBE (89 mM Tris-borate; 89 mM boric acid and 2mM EDTA). Electrophoresis was carried out at voltages and times depending on the size and concentration of the gel and the fragment separation required. DNA size markers used to determine fragment sizes were ϕX 174 Hae III or λ DNA digested with Hind III. The migration of DNA was visualised by exposure to short wave UV irradiation from a UV transilluminator and the image photographed using a Polaroid camera.

A3.7 Alkaline gel electrophoresis

The quality of cDNA was assessed on a 1% alkaline gel (1mM EDTA, pH 8; 250 mM NaOH) along with λ DNA digested with Hind III marker end labelled with γ ^{32}P dATP. 2 µl of the labelled cDNA sample was loaded onto the gel along with 8 µl alkaline gel loading buffer (0.3 M NaOH; 6 mM EDTA, pH 8.0; 18% Ficoll, 0.15% Bromophenol Blue, 0.25% Xylene Cyanol). The gel was run at 15-20 V over a period of 3-4 hours. The cDNA was fixed by soaking the gel in 7% TCA solution overnight. The gel was then dried and subjected to autoradiography.

A3.8 Quantification of DNA and RNA

The concentration of DNA, RNA and oligonucleotides was determined by measuring absorbance at wavelengths of 260 and 280 nm. The ratio between the readings at 260 nm and 280 nm provided an estimation of purity. Pure preparations of DNA and RNA have OD_{260}/OD_{280} values of 1.8 and 2.0 respectively. An OD of 1 unit at 260 nm was assumed to be equivalent to 50 µg/ml for double stranded DNA,

40 μg/ml for RNA and 37 μg/ml for oligonucleotides. The concentration of small quantities of DNA was estimated by running samples on an agarose gel and comparing the intensity of ethidium bromide fluorescence in the sample with known amounts of bacteriophage λ DNA digested with Hind III.

A3.9 Preparation of competent cells (strains : W3110, BMH 71.18, TG1)

Method 1

A 20 ml overnight culture of the strain was prepared and 100 μl was used to inoculate 20 mls of LB medium. The culture was grown at 37°C, 250 rpm, until an OD_{550nm} of about 0.3 units was reached. The cells were cooled briefly on ice before pelleting at 3000-4000 rpm at 4°C. The cells were resuspended gently in 10 mls of sterile ice cold 0.1M $CaCl_2$ and incubated on ice for 1 hour. The cells were pelleted and resuspended in 2 ml of 0.1M $CaCl_2$ solution prior to transformation.

For transformation of M13mp18 into TG1, the competent cells were prepared by inoculating 20 mls of 1 × TY with 100 μl of an overnight culture grown from a colony selected on minimal agar.

Method 2

A more efficient method for the preparation of competent cells was adopted. 200 ml of 1 X TY was inoculated with 200 μl of an overnight culture and incubated with shaking at 37°C to an OD of 0.3-0.5 measured at 550 nm. After incubation on ice for 30 minutes the culture was centrifuged at 2500 rpm for 5 minutes at 4°C. The pellet was resuspended in 100 ml of ice cold 100mM $MgCl_2$,

2mM Tris HCl (pH 7.4). After centrifugation at 2500 rpm for 5 minutes at 4°C, the cell pellet was resuspended in 100 ml of ice cold 100 mM $CaCl_2$,2 mM Tris HCl (pH 7.4). After incubation on ice for 60 minutes, the cells were pelleted and gently resuspended in 10 ml of 100 mM $CaCl_2$, 2mM Tris HCl (pH 7.4), 15% glycerol. The competent cells were aliquoted in 200-400 µl volume into microfuge tubes and either frozen at -70°C or used immediately.

A3.10 Transformation of plasmid DNA

200 µl aliquots of competent cells were mixed with the transforming DNA (10 ng) in a pre-cooled 15 ml polypropylene tube (Falcon). The mixture was incubated on ice for 1 hour before incubating at 42°C for 45 seconds (heat shocking). After 2 minutes on ice, the volume was increased to 1 ml with LB media and incubated at 37°C, 250 rpm for 1 hour. 200 µl of the transformed mixture was plated onto LB media plates containing the appropriate antibiotic and incubated inverted at 37°C.

A3.11 Preparation of single stranded template

Infected cells from recombinant colourless plaques were grown to produce single stranded template for the sequencing reaction. 1 ml of an overnight culture of *E.coli* cells was used to inoculate 100 ml 1 X TY. 1.5 ml aliquots of the culture were then dispensed into sterile 10 ml culture tubes. A colourless plaque was inoculated into each tube. The cells were pelleted by centrifugation at 13000 rpm for 5 minutes after incubation at 37°C for 5 hours. The supernatant was then

recentrifuged to ensure that all cells were removed. After the addition of 4% PEG 6000 and 0.5 M NaCl themixture was incubated at RT. for 15 minutes. The supernatant was discarded after centrifugation at 13000 rpm for 5 minutes. All the remaining traces of PEG were removed with a tissue after a subsequent centrifugation at 13000 rpm for two minutes. The viral pellet was resuspended in 100 µl TE buffer. It was then extracted with an equal volume of PhOH followed by an extraction with $CHCl_3$. DNA was precipitated by the addition of 0.1 volume 3M sodium acetate, pH 5.2 and 2.5 volumes EtOH followed by incubation at -20°C overnight. The pellet was then washed with 1 ml ice cold EtOH and resuspended in 50 µl TE buffer.

A3.12 SDS-Polyacrylamide gel Electrophoresis of proteins

Proteins were analysed by SDS-PAGE in 1.5 mm thick, 12.5 cm long slab gels with an upper 'stacking' gel and a lower 'separating' gel. Electrophoresis was carried out with a discontinuous buffer system in which the buffer in the reservoirs was of a different pH and ionic strength from the buffer used to cast the gel. The sample and the stacking gel contain Tris HCl (pH 6.8), the upper and lower buffer reservoirs contain Tris glycine (pH 8.3) while the resolving gel contains Tris HCl (pH 8.8).

The separating gel was prepared by mixing the following solutions: 14 ml 40% (w/v) acrylamide stock solution [acrylamide: N, N' methylene bis acrylamide; 19:1]; 9.4 ml pH 8.8 buffer (18.2 g% Tris HCl, 0.23 g% EDTA and 0.4 g% SDS) and 14.1 ml water. After the addition of 20 µl TEMED and 160 µl 25% ammonium

persulphate the separating gel was immediately poured into the gap between 2 glass plates. This solution was overlaid with butanol till polymerisation was complete. The overlay was then tipped off and the top of the gel rinsed several times with Milli Q water. The stacking gel was then prepared by mixing the following solutions: 3.6 ml 40% (w/v) acrylamide stock solution [acrylamide: N, N' methylene bis acrylamide; 19:1] solution; 6 ml pH 6.8 buffer (3g% Tris HCl, 0.12 g% EDTA, 0.2 g% SDS) and 14 ml water. After the addition of 12 µl TEMED and 300 µl 25% ammonium persulphate the stacking gel was poured onto the surface of the polymerised separating gel and a Teflon comb was immediately inserted into the stacking gel solution carefully. The gel was allowed to set in a vertical position at room temperature for at least 40 minutes

Samples were taken up in 50 mM Tris HCl, pH 6.8; 25% (v/v) glycerol; 0.1% (v/v) bromophenol blue; 6% (w/v) SDS. In the case of reducing SDS-PAGE, 4% (v/v) β- mercapto ethanol was included in the loading buffer. The PAGE running buffer used in both electrode compartments contained 0.025 M Tris HCl, 0.192 M glycine and 0.1% w/v SDS (pH 8.3). The gels were run at 50V overnight. Proteins were stained by gently shaking the gel in 0.1% (w/v) Coomassie Brilliant Blue R; 50% (v/v) methanol and 10% (v/v) acetic acid and subsequently destained with 10% acetic acid, 10% methanol.

A3.13 DNA Sequencing

DNA sequencing was performed according to Sanger's dideoxy chain termination method which involves the synthesis of a DNA strand in vitro using a

DNA polymerase with a single stranded DNA template.

For every set of four sequencing lanes a single annealing and subsequent labelling reaction was used. In a micro centrifuge tube 1 μg of single stranded DNA template was taken in the presence of 40 mM Tris HCl pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl and 0.5 pmole -40 primer. The total volume was made upto 10 μl. It was then warmed to 65°C and allowed to cool to room temperature. After the addition of 6.3 mM DTT, labelling mixture (1.5 μM dGTP, 1.5 μM dCTP, 1.5 μM dTTP), 5 μCi [^{35}S] dATP and Sequenase Version 2.0 enzyme, the annealed template primer was incubated at room temperature for 5 minutes. 3.5 μl of this mixture was then added to 2.5 μl of ddGTP, ddCTP, ddTTP and ddATP termination mixtures respectively. This mixture was then warmed at 37°C for 5 minutes. After the addition of 4 μl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) the mixture was stored at -20°C until further use.

Sequencing reactions were analysed by polyacrylamide gel electrophoresis on 0.4 mm 6% denaturing polyacrylamide gels. Gels were prepared by dissolving 31.5 g urea in a solution containing 15 ml 40% (w/v) acrylamide stock (acrylamide : N,N' methylenebisacrylamide; 19:1), 7.5 ml 10 × TBE and 450 μl 10% ammonium persulphate made upto a volume of 75 ml with water. Gels were poured after the addition of 45 μl TEMED. Gels were run on a gel apparatus in TBE buffer.Gels were pre electrophoresed at 40 W for 1 hour to remove the unpolymerised acrylamide monomer and to pre warm the gel to its optimum temperature of 55°C. The sequencing reactions were denatured by heating at 75°C for 3 minutes before

loading onto the gel. For each sequencing reaction 2-3 µl of each sample was loaded per well and run at a constant power of 40 W. Electrophoresis was continued until the bromophenol blue dye reached the bottom of the gel for short runs, or twice the length of the gel for long runs. After electrophoresis the gels were washed for 20- 30 minutes with a solution of 10% v/v methanol, 10% v/v acetic acid to fix DNA and to leach urea. The gel was then transferred onto Whatmann 3 MM paper and dried under vacuum at 80°C on a slab drier for 1 hour. Autoradiography was performed overnight without an intensifying screen.

References

Abraham, N., Miceli, M.C., Parnes, J.R. and Veillette, A. (1991) *Nature* **350**, 62-66.

Acha-Orbea H., Mitchell, D. J. and Timmermann, L (1988) *Cell* **54**, 263-273.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) *J. Mol. Biol.* **215**, 403-410.

Alzari, P.M., Lascombe, M.B. and Poljak, R.J. (1988) *Ann. Rev. Immunol.* **6**, 555-580.

Amit, A.G., Mariuzza, R.A., Phillips, S.E.V. and Poljak, R.J. (1986) *Science* **233**, 747-753.

Aqvist, J., van Gunsteren, W.F., Leifonmark, M. and Tapia, O. (1985) *J. Mol. Biol.* **183**, 461-477.

Artimiuk, P.J. and Blake, C.C.F. (1981) *J. Mol. Biol.* **152**, 737-762.

Baccala, R., Kono, D.H., Walker, S., Balderas, R.S. and Theofilopoulos, A.N. (1991) *Proc. Natl. Acad. Sci USA* **88**, 2908-2912.

Baniyash, M., Garcia-Morales, P., Bonifacino, J.S., Samelson, L.E. and Klausner, R.D. (1988) *J.Biol. Chem.* **263**, 9874-9878.

Baniyash, M., Hsu, V.W., Seldin, M.F. and Klausner, R.D. (1989) *J. Biol. Chem.* **264**, 13252-13257.

Bank, I., DePinho R.A., Brenner, M.B., Cassimeris, J., Alt, F.W. and Chess, L. (1986) *Nature* **322**, 179-181.

Benner, S.A. (1989) *Advan. Enzyme Regul.* **28**, 219-236.

Bentley, G.A., Boulot, G., Riottot, M.M. and Poljak, R.J. (1990) *Nature* **348**, 254-257.

Bentley, G.A., Boulot, G., Karjalainen, K. and Mariuzza, R.A. (1995) *Science* **267**, 1984-1987.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535-542.

Bevan, M.J. (1984) *Immunol. Today* **5**, 128.

Bhat, T.N., Bentley G.A., Fischmann, T.O., Boulot, G. and Poljak, R.J. (1990) *Nature* **347**, 483-484.

Bigby, M., Markowitz, J.S., Bleicher, P.A., Grusby, M.J., Simha, S., Siebrecht, M., Wagner, M., Nagler-Anderson, C. and Glimcher, L.H. (1993) *J. Immunol.* **151**, 4465-4475.

Bill, J., Yague, J., Appel, V.B., White, J., Horne, G., Erlich, H.A. and Palmer, E. (1989) *J. Exp. Med.* **169**, 115-133.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennet, W.S., Strominger, J.L. and Wiley, D.C. (1987) *Nature* **329**, 506-512.

Blake, C.C.F., Koenig, D.F., Mair G.G., North, A.C.T., Phillips, D.C. and Sarma V.R. (1965) *Nature* **206**, 757-761.

Bollum, F.J. and Chang L.M. (1981) *J. Biol. Chem.* **256**, 8767-8770.

Bott, R. and Sarma, R. (1976) *J. Mol. Biol.* **106**, 1037-1046.

Braciale, T., Andrew, M. and Braciale, V. (1981) *J. Exp. Med.* **153**, 1371.

Braden. B.C., Fields, B.A., Ysern, X., Goldbaum, F.A., Dall'Acqua, W., Schwarz,

F.P., Poljak, R.J. and Mariuzza., R.A. (1996) *J. Mol. Biol.* ***257,*** 889-894.

Brady, R.L., Dodson, E.J., Dodson, G.G., Lange, G., Davis, S.J., Williams, A.F. and Barclay, A.N. (1993) *Science* **260**, 979-983.

Brenner, M.B., Mclean, J., Dialynas, D., Strominger, J.L., Smith, J.A. and Owen, F.L. (1986) *Nature* **322**, 145-149.

Broeren, C.P., Verjans, G.M., Van Eden, W., Kusters, J.G., Lenstra, J.A. and Logtenberg, T. (1991) *Eur. J. Immunol.* **21**, 569-575.

Brooks, B., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. (1983) *J. Comp. Chem.* **4,** 187-217.

Bruccoleri, R.E. and Karplus, M. (1987) *Biopolymers* **26**, 137-168.

Buck, L. and Axel, R. (1991) *Cell* **65**, 175

Bucy, P., Chen, C.L., Cihak, J., Löaxh, U. and Cooper, M. (1988) *J. Immunol* **141,** 2200-2205.

Caccia, N., Bruns, G.A., Kirsch, I.R., Hollis, G.F., Bertness, V. and Mak, T.W. (1985) *J. Exp. Med.* **161,** 1255-1260.

Cammarota, G., Schierle, A., Takacs, B., Doran, D., Knorr, R., Bannworth, W., Guardiola, J. and Sinigaglia, F. (1992) *Nature* **356**, 799-801.

Chan, A., Iwashima, M., Turck, C. and Weiss, A. (1992) *Cell* **71**, 649-642

Chan, A.C., Desai, D.M. and Weiss, A. (1994) *Annu. Rev. Immunol.* **12**, 555-592.

Chicz, R.M., Urban, R.G., Lane, W.S., Gorga, J.C., Stern, L.J., Vignali, D.A. and Strominger, J.L. (1992) *Nature* **358,** 764-768.

Chien, Y.H., Iwashima, M., Kaplan, K.B., Elliott, J.F. and Davis, M.M. (1987) *Nature* **327,** 677-682.

Choi, Y., Kotzin, B., Herron, L., Callahan, J., Marrack, P. and Kappler, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8941-8945.

Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156.

Chothia, C., Lesk A. M., Tramontano, A., Levitt, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M. and Poljak, R.J. (1989) *Nature* **342**, 877-883.

Chothia, C. and Lesk, A.M. (1987) *J. Mol. Biol.* **196**, 901-917.

Chothia, C., Boswell, D.R. and Lesk, A.M. (1988) *EMBO J.* 7, 3745-3755.

Claverie, J.M., Prochnicka-Chalufour, A. and Bougueleret, L. (1989) *Immunol. Today* **10**, 10-14 .

Clayberger, C., Parham, P., Rothbard, J., Ludwig, D.S., Schoolnik, G.K. and Krensky, A.M. (1987) *Nature* **330**, 763-765.

Collins, M.K.L., Goodfellow, P.N., Spurr, N.K., Solomon, E., Tanigawa, G., Tonegawa, S. and Owen, M.J. (1985) *Nature* **314,** 273-274.

Coppin, H.L., Carmichael, P., Lombardi, G., L'Faqihi, F.E., Salter, R., Parham, P., Lechler, R.I. and de Preval, C. (1993) *Eur. J. Immunol.* **21**, 569-575.

Correa, I., Bix, M., Liao, N.S., Zijlstra, M. and Jaenisch, R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 653-657.

Cresswell, P. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8188-8192.

Cresswell, P. (1994) *Annu. Rev. Immunol.* **12**, 259-293.

Cresswell, P., Blum, J.S., Kelner, D.N. and Marks, M.S. (1987) *CRC Crit. Rev. Immunol.* 7, 31-53.

Danska, J.S., Livingstone, A.M., Paragas, V., Ishihara, T. and Fathman, C.G.

(1990) *J. Exp. Med.* **172**, 27-33.

Darsley, M.J. and Rees, A.R. (1985*a*) *EMBO J.* **4**, 383-392.

Darsley, M.J. and Rees, A.R. (1985*b*) *EMBO J.* **4**, 393-398.

Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) *In Atlas of protein sequence and structure* (ed. Dayhoff, M.O), Vol **5**, suppl.3, p.345, The National Biomedical Research Foundation, Washington, D.C.

Davis, M.M. and Bjorkman, P.J. (1988) *Nature* (London) **334**, 395-402.

de la Paz, P., Sutton, B., Darsley, M.J., Rees, A.R. (1986) *EMBO J.* **5**, 415-425.

De Libero. G., Casorati, G., Giachino, C., Carbonara, C., Migone, N., Matzinger, P. and Lanzavecchia, A. (1991) *J. Exp. Med.* **173**, 1311- 1322.

Doyle, C. and Strominger, J.L. (1987) *Nature* (London) **330**, 256-259.

Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) *Cell* **19**, 981-992.

Eckels, D. D., Gorski, J., Rothbard, J. and Lamb, J.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8191.

Ehrich, E. W., Devaux, B., Rock, E.P., Jorgensen, J.L., Davis, M.M. and Chien, Y. (1993) *J. Exp. Med.* **178**, 713-722.

Epp, O., Lattman, E.E., Schiffer, M., Huber, R. and Palm, W. (1975) *Biochemistry* **14**, 4963-4975.

Falk, K., Rotzschke, O., Stevanovic, S., Jung, G. and Rammensee, H.G. (1991) *Nature* 351, 290-296.

Feng and Doolittle (1987) *Journal of Molecular Evolution* **25**, 351-360.

Fields, B.A., Ober, B., Malchiodi, E.L., Lebedeva, M.I., Ysern, X., Ward, E.S. and

Mariuzza, R.A. (1995) *Science* **270**, 1821-1824.

Fink, P.J., Matis, L.A., McElligott, D.L., Bookman, M. and Hedrick, S.M. (1986) *Nature* **321**, 219-226.

Fleury S., Croteau G. and Sekaly, R.P. (1991) *Semin. Immunol.* **3,** 177.

Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A. and Wilson, I.A. (1992) *Science* **257**, 919-927.

Genevée, C., Chung, V., Diu, A., Hercend, T. and Triebel, F. (1994) *Mol. Immunol.* **31**, 109-115.

Glockshuber, R., Malia, M., Pfitzinger, I. and Pluckthun, A. (1992) *Biochemistry* **31**, 1270-1279.

Goronzy, J.J., Xie, C., Hu, W., Lundy, S.K. and Weyand, C.M., (1993) *J Immunol.* **151**, 825-836.

Grusby, M.J., Auchincloss, H., Lee, R., Johnson, R.S., Spencer, J.P., Zijlstra, M., Jaenisch, R., Papaioannou, V. and Glimcher L. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3913-3917.

Gribskov, M., McLachlan, A.D. and Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4355-4358.

Harper, M., Lema, F., Boulot, G. and Poljak, R.J. (1987) *Mol. Immunol.* **24**, 97-108.

Hata, S., Bremner, M.B. and Krangel, M.S. (1987) *Science* **238**, 678-682.

Heath, W.R., Hurd, M.E., Carbone, F.R. and Sherman, L.A. (1989) *Nature* **341**, 749.

Heber-Katz, E. and Acha-Orbea, H (1989) *Immunol. Today* **10**, 164-169.

Hedrick, S.M., Engel, I., McElligott, D.L., Fink, P.J., Hsu, M-L., Hansburg, D., Matis, L.A. (1988) *Science* **239**, 1541.

Henikoff and Henikoff (1992) *Proc Natl Acad Sci. USA* **89**, 10915-10919

Higgins and Sharp (1989) *CABIOS* **5**, 151-153.

Hilyard, K. (1991) *D.Phil. thesis*, Oxford.

Hiromatsu, K., Yoshikai, Y., Matsuzaki, G., Ohga, S., Muramori, K., Matsumoto, K., Bluestone, J.A., Nomoto, K., (1992) *J. Exp. Med.* **175**, 49-56.

Hochstenbach, F., Parker, C., McLean, J., Gieselmann, V., Band, H., Bank, I., Chess, L., Spits, H., Strominger, J.L. and Siedman, J.G. (1988) *J. Exp. Med.* **168**, 761-776.

Hockett, R.D., Villartay, J., Pollock, K., Poplack, D.G., Cohen, D.I. and Korsmeyer, S.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9694-9698.

Hong, S.C., Chelouche, A., Lin, R., Shaywitz, D., Braunstein, N.S. and Glimcher, L. (1992) *Cell* **69**, 999-1009.

Hunt, D.F., Henderson, R.A., Shabanowitz, J., Cox, A.L. Sakaguchi, K., Michel, H., Sevilir, N. (1992*a*) *Science* **255**, 1261-1266.

Hunt, D.F., Hanspeter, M., Dickinson, T.A., Shabanowitz, J., Cox, A.L. Sakaguchi, K. (1992*b*) *Science* **256**, 1817-1820.

Hurley, C.K., Steiner, N., Wagner, A., Geiger, M.J., Eckels, D.D. and Rosen-Bronson, S. (1993) *J. Immunol* **150**, 1314-1324.

Itohara, S., Nakanishi, N., Kanagawa, O., Kubo, R. and Tonegawa, S. (1989) *Proc. Natl. Acad. Sci. USA* **86**: 5094-5098.

Janeway, C.A. (1992) *Annu. Rev. Immunol.* **10**, 645-674.

Jardetsky, T.S., Lane, W.S., Robinson, R.A., Madden, D.R. and Wiley, D.C. (1991) *Nature* **353**, 326-329.

Jeffrey, P. (1989) *D.Phil. thesis*, Oxford.

Jerne, N.K., (1971) *Eur. J. Immunol.* **1**, 1.

Jin Y.J., Clayton, L.K., Howard, F.D., Koyasu, S., Sieh, M., Steinbrich R., Tarr, G.E. and Reinherz, E.L. (1990) *Proc Natl Acad Sci. USA* **87**, 3319-3323.

Jores, R., Alzari, P.M. and Meo, T., (1990) *Proc Natl Acad Sci. USA* **87**, 9138-9142.

Jorgensen, J.L., Esser, U., Fazekas, de St. Groth B., Reay, P.A. and Davis, M.M. (1992) *Nature*, **355**, 224-230.

Kabat, E.A., Wu, T.T. and Bilofsky, H. (1977) *J. Biol. Chem.* **252**, 6609-6616.

Kabat, E.A., Schiffer, M. and Wu, T.T. (1992) *Immunogenetics* **35**, 224-234.

Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M. and Gottesman, K.S. (1987) *Sequences of proteins of Immunological Interest*, Bethesda, Md; Natl. Inst. Health; 4th Edition.

Kallenbach, S., Doyen, N., D' Andon, M.F. and Rougeon, F. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2799.

Kaufmann, S.H.E. (1994) *Curr. Opin. Immunol.* **6**, 518-525.

Kay, R.A., Snowden, N., Hajeer, A.H., Boylston, A.W. and Ollier, W.E. (1994) *Eur. J. Immunol.* **24**, 2863-2867.

Kaye, J. and Hedrick, S.M. (1988) *Nature* **336**, 580-583.

Klein, J. (1986) *Natural History of the Major Histocompatibility Complex.* John

Wiley and Sons, New York.

Konig R., Huang, L.Y. and Germain, R. (1992) *Nature* **356,** 796-798.

Koning, F., Stingl, G., Yokoyama, W.M., Yamada, H., Maloy, W.L., Tschacler, E., Shevach, E.M. and Coligan, J.E. (1987) *Science* **236**, 834-837.

Krangel, M.S., Band, H., Hata, S., McLean, J. and Brenner, M.B. (1987) *Science* **237,** 64-67.

Kuziel, W.A., Takashima, A., Bonyhadi, M., Bergstresser, J., Allison, J.P. and Tigelaar, R.E. (1987) *Nature* **328,** 263-266.

Kvist, S., Wiman, K., Claesson, L., Peterson, P.A. and Dobberstein, B. (1982) *Cell* **29,** 61- 69.

Ladel, C.H., Flesch, I.E., Arnoldi, J., Kaufmann, S.H. (1994) *J. Immunol.* **153,** 3116-3122.

Lafaille, J.J., DeCloux, A., Bonneville, M., Takagaki, Y. and Tonegawa, S. (1989) *Cell* **59,** 859-870.

Lai, M.Z., Huang, S.Y., Briner, T.J., Guillet, J.G., Smith, J.A. and Gefter, M.L. (1988) *J.Exp. Med.* **168,** 1081-1097.

Lai, C. and Lemke, G. (1991) *Neuron* **6**, 691-704.

Lamb, C.A., Yewdell, J.W., Bennink, J.R. and Cresswell, P. (1991) *Proc Natl Acad Sci. USA* **88,** 5998-6002.

Landau, N.R., Schatz, D., Rosa, M. and Baltimore, D. (1987) *Mol. Cell. Biol.* **7**, 3237.

Lanier, L.L., Federspiel, N.A., Ruitenberg, J.J., Phillips, J.H., Allison, J.P., Littman, D. and Weiss, A. (1987) *J. Exp. Med.*, **165,** 1076-1094.

Larrick, J.W., Danielsson, L., Brennen, C.A., Abrahamson, M., Fry, K.E. and Borrebaeck C.A. (1989) *Biochem. Biophys. Res. Commun.* **160**, 1250.

Lesk, A.M. and Chothia C. (1982) *J. Mol. Biol.* **160**, 325-342.

Lesk, A.M. and Chothia C. (1988) *Nature* **335**, 188-190.

Li, Y., Szabo, P. and Posnett, D.N. (1988) *J. Immunol.* **140,** 1300-1303.

Lieber, M.R. (1992) *Cell* **70**, 873-876.

Loh, E.Y., Lanier, L.L., Turck, C.W., Littman, D.R., Davis, M.M., Chien, Y.H. and Weiss, A., (1987) *Nature* **330**, 569-572.

Loh, E.Y., Elliot, J.F., Cwirla, S., Lanier, L.L. and Davis, M.M., (1989) *Science*, **243**, 217-220.

Machamer, C.E. and Cresswell, P. (1982) *J. Immunol.* **129,** 2564-2569.

Madden, D.R., Gorga, J.C., Strominger, J.L. and Wiley, D.C. (1991) *Nature* **353**, 321-325.

Madden, D.R., Gorga, J.C., Strominger, J.L. and Wiley, D.C. (1992) *Cell* **70**, 1035-1048.

Madden, D.R., Garboczi, D.N. and Wiley, D.C., (1993) *Cell* **75**, 693-708.

Male, D., Cooke, A., Owen, M., Trowsdale, J. and Champion, B. (Eds) *Advanced Immunology,* 1996.

Malissen, M., Trucy, J., Letourneur, F., Rebai, N., Dunn, D.E., Fitch, F.W., Hood, L. and Malissen, B. (1988) *Cell* **55**, 49-59.

Marguerie, C., Lunardi, C. and So, A. (1992) *Immunol. Today* **13**, 336-338.

Marston, A.O. (1986) *Biochem. J.* **240**, 1-12.

Martin, A.C.R., Cheetham, J.C. and Rees, A.R. (1989) *Proc Natl Acad Sci. USA*

86, 9268-9272.

Matis, L.A., Sorger, S.B., McElligott, D.L., Fink, P.J. and Hedrick, S.M. (1987) *Cell* **51,** 59-69.

Matzinger, P. and Bevan, M.J. (1977) *Cell Immunol.* **29**, 1-5.

Mombaerts, P., Arnoldi, J., Russ, F., Tonegawa, S., Kaufmann, S. (1993) *Nature* (London) **365**, 53-56.

Moonka, D. and Loh, E.Y. (1994) *J. Immunol. Methods* **169,** 41-51.

Morel, P., Livingstone, A.M. and Fathman, C.G. (1987) *J. Exp. Med.* **166**, 583-588.

Moss, P.A., Rosenberg, W.N. and Bell, J.I. (1992) *Annu. Rev. Immunol.* **10**, 71-96.

Murre, C., Quertermous, T., Strauss, W., Dialynas, D.P., Strominger, J.L. and Seidman, J.G., (1986) *Nature* **322**, 184-187.

Neefjes, J.J., Stollorz, V., Peters, P.J., Geuze, H.J. and Ploegh, H.L. (1990) *Cell* **61**, 171-183.

Newcomb, J.R. and Cresswell, P. (1993) *J. Immunol.* **150**, 499-507.

Norment, A.M., Salter, R.D., Parham, P., Engellhard, V.H. and Littman, D.R. (1988) *Nature* **336**, 79-81.

Novotny, J., Bruccoleri, R.E., Newell, J., Murphy, D., Haber, E. and Karplus, M., (1983) *J. Biol. Chem.* **258**, 14433-14437.

Novotny, J. and Haber, E. (1985) *Proc Natl Acad Sci. USA* **82,** 4592-4596.

Offner, H., Hashim, G.A. and Vandenbark, A.A. (1991) *Science* **251**, 430-432.

Orloff, D.G., Frank, S.F., Robey, F.A., Weissman, A.M. and Klausner, R.D.(1989) *J. Biol. Chem.* **264,** 14812-14817.

Orloff, D.G., Ra, C., Frank, S.J. and Kinet, J.P. (1990) *Nature* **347,** 189-191.

Padlan, E.A., Silverton, E.W., Sheriff, S., Cohen, G.H., Smith-Gill, S.J. and Davies, D.R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5938-5942.

Padovan, E., Casorati, G., Dellabona, P., Giachino, C., Lanzavecchia, A. (1995) *Ann. N Y Acad. Sci.* **756**, 66-70.

Paliard, X., Kappler, J.W., Marrack, P. and Kotzin, B.L. (1991) *Science* **253,** 325-329.

Parham, P., Clayberger, C., Zorn, S.L., Ludwig, D.S., Schoolnik, G.K. and Krensky, A.M. (1987) *Nature* **325**, 625-628.

Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448.

Phillips, D.C., Acharaya, K.R., Handoll, H.H., Stuart, D.I. (1987) *Biochem. Soc. Trans.* **15,** 737-744.

Poljak, R.J., Amzel, L.M., Avey, H.P., Chen, B.L., Phizackerley, R.P., Saul, F. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3305-3310.

Raulet, D.H. (1994) *Curr. Biol.* **4**, 246-248.

Roberts, S. (1987) *D.Phil. thesis*, Oxford.

Rock, E.P., Sibbald, P.R., Davis, M.M., Chien, Y. (1994) *J. Exp. Med.* **179**, 985-989.

Rock, K. (1996) *Immunol. Today* **17**, 131-137

Romero, P., Corradin, G., Luescher, I.F., Maryanski, J.L. (1991) *J. Exp. Med.* **174,** 603- 612.

Rotzschke, O., Falk, K., Wallny, H.J., Faath, S., Rammensee, H.G. (1990) *Science* **249,** 283-287.

Rowen, L., Koop, B.F., Hood, L. (1996) *Science* **272,** 1755-1762.

Rudensky, A.Y., Preston-Hurlburt, P., al-Ramadi, B.K., Rothbard, J., Janeway, C.A. Jr. (1992) *Nature* **359,** 429-431.

Ryu, S.E., Kwong, P.D., Trunch, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.H., Axel, R., Sweet, R.W., Hendrickson, W.A. (1990) *Nature* **348,** 419-426.

Salter, R.D., Benjamin, R.J., Weseley, P.K., Baxton, S.E., Garrett, T.P.J., Clayberger, C., Krensky, A.M., Normant, A.M., Littman, D.R., Parham, P. (1990) *Nature* **345,** 41-46.

Samelson, L.E., Phillips, A.F., Luong, E.T., Klausner, R.D., (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4358-4362.

Samelson, L.E. and Klausner, R.D., (1992) *J. Biol. Chem.* **267,** 24913-24916.

Sanders, S., Fox, R., Kavathas, P., (1991) *J. Exp. Med.* **174**, 371-379.

Schild, H., Mavaddat, N., Litzenberg, C., Ehrich, E., Davis, M.M., Bluestone, J.A., Matis, L., Draper, R., Chien, Y.H. (1994) *Cell* **76**, 17-27.

Schwartz, R.H. (1985) *Annu. Rev. Immunol.* **3**, 237-261.

Searle, S.M.J. and Rees, A.R. (1996) Manuscript in preparation.

Segurado., O.G., and Schendel, D.J. (1993) *Electrophoresis* **14,** 747-752.

Sheriff, S., Silverton, E.W., Padlan, E.A., Cohen, G.H., Smith-Gill, S.J., Finzel, B.C., Davies D.R., (1987) *Proc. Natl. Acad. Sci. USA* **84**, 8075-8079.

Silver, M.L., Guo, H.C. and Wiley, D.C., (1992) *Nature* **360**, 367-369.

Smith and Waterman, (1981) *Advances in Applied Mathematics* **2**, 482-489.

Smith-Gill, S.J., Lavoie, T.B., Mainhart, C.R., (1984) *J. Immunol.* **133**, 384-393.

Sneath, P.H.A., Sokal, R.R. (1973) *Numerical Taxonomy,* 230-234, W.H. Freeman

and Company, San Francisco, California, USA.

Sottini, A., Imberti, L., Gorla, R., Cattaneo, R. and Primi, D. (1991) *Eur. J. Immunol.* **21,** 461-466.

Sredni, B. and Schwartz, R.H. (1980) *Nature* **287,** 855-857.

Stingl, G. (1987*a*) *Proc. Natl. Acad. Sci. USA* **84,** 2430-2434.

Stingl, G., Koning, F., Yamada, H., Yokoyama, W.M., Tschachler, E., Bluestone, J.A. (1987*b*) *Proc. Natl. Acad. Sci. USA* **84,** 4586-4589.

Stern, L.J., Brown, J.H., Jardetzky, T.S., Gorga, J.C., Urban, R.G., Strominger, J.L., Wiley, D.C. (1994) *Nature* **368,** 215-221.

Stern, L.J. and Wiley, D.C. (1994) *Structure* **2,** 245-251.

Sumida, T., Yonaha, F., Maeda, T., Tenabe, E., Koike, T., Tomioka, H., Yoshida, S. (1992) *J. Clin. Invest.* **89,** 681-685.

Tan, K. (1988) *Cell* **54,** 247-261.

Townsend, A., Ohlen, C., Bastin, J., Ljungren, H.G., Foster, L., Karre, K. (1989) *Nature* (London) **340,** 443-448.

Toyonaga, B., Yoshikai, Y., Vadasz, V., Chin, B. and Mak, T.W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8624-8628.

Tramontano, A. and Schloeder, D. (1989) *Methods Enzymol* **178,** 531-550.

Tsuji, M., Mombaerts, P., Lefrancois, L., Nussenzweig, R.S., Zavala, F. and Tonegawa S. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 345-349.

Uematsu, Y. (1991) *Immunogenetics,* **34,** 174-178.

Van Bleek, G.M. and Nathenson, S.G. (1990) *Nature* **348,** 213-216.

Wang, J., Yan, Y., Garrett, T.P.J., Liu, J., Rodgers, D.W., Garlick, R.L., Tarr,

G.E., Husain, Y., Reinherz, E.L. and Harrison, S.C. (1990) *Nature* **348,** 411-418.

Wedderburn, L.R., Searle, S.M.J., Rees, A.R., Lamb, J.R. and Owen, M.J. (1995) *Eur. J. Immunol.* **25,** 1654-1662.

Weiss, A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6998-7002.

Weissman, A.M., Baniyash, M., Hou, D., Samelson, L.E., Burgess, W.H. and Klausner, R.D. (1988) *Science* **239,** 1018-1021.

White, J., Herrman, A., Pullen, A.M., Kubo, R.T., Kappler, J.W. and Marrack, P. (1989) *Cell* **56**, 27-35.

Wilbur and Lipman (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726-730

Winoto, A., Chung, S., Abraham, J. and Echols, H. (1986) *Nature* **324,** 679-682.

Wu, T.T. and Kabat, E.A. (1970) *J. Exp. Med.* **132**, 211-250.

Yamanaka, K., Kwok, W.W., Mickelson, E.M., Masewicz, S., Smith, F. and Nepom, G.T. (1993) *Transplantation* **55**, 1167-1175.

Yoshikai, Y., Clark, S.P., Taylor, S., Sohn, U., Wilson, B.I., Minden, M.D. and Mak, T.W. (1985) Nature 316, 837-840.

Young, A.C.M., Zhang, W., Sacchettini, J.C. and Nathenson, S.G. (1994) *Cell* **76,** 39-50.

Zeliszewski, D., Golvano, J.J., Gaudebout, P., Dorval, I., Freidel, C., Gebuhrer, L., Betuel, H., Borras-Cuesta, F. and Sterkers, G. (1993) *J. Immunol.* **151,** 6237-6247.

Zhang, W., Young, A.C.M., Imarai, M., Nathenson, S.G. and Sacchettini, J.C. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8403-8407.